HUMAN FACTORS IN FIELD TESTING

FINAL REPORT

LS-FR-71-1

W. G. MATHENY
G. W. PATTERSON, JR.
G. I. EVANS

PREPARED FOR

ENGINEERING PSYCHOLOGY PROGRAMS
PSYCHOLOGICAL SCIENCES DIVISION
OFFICE OF NAVAL RESEARCH

CONTRACT N00014-67-C-0315

WORK UNIT NO. NR 196-067

DECEMBER 1971

Approved for public release; distribution unlimited.

LIFE SCIENCES, INC.

86

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Life Sciences, Inc. 227 Loop 820 Northeast Hurst, Texas 76053	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE

HUMAN FACTORS IN FIELD TESTING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report - LS-FR-71-1

5. AUTHOR(S) (First name, middle initial, last name)

W. G. Matneny, G. W. Patterson, Jr. and G. I. Evans

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1971	78	100

8a. CONTRACT OR GRANT NO. N00014-67-C-0315 b. PROJECT NO. Work unit No. NR 196-067 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Engineering Psychology Programs Psychological Sciences Division Office of Naval Research, Arlington, Va.

13. ABSTRACT

The primary aim of this study has been the development and test of methods for evaluating operator performance during field test of complex man-machine systems. Recommendations are presented which are aimed at strengthening the human factors evaluation process in both the field evaluation and the development phases of a system. However, during the study field evaluation took precedence with respect to development and trial of test procedures and techniques.

Recommendations are made based upon the study of the development and field test of specific Naval aircraft systems. Field evaluation techniques and methods were tested during the evaluation of one of these systems - the P-3C anti-submarine warfare system. These recommendations deal with the identification of measurement points, measurement scales, evaluation criteria and levels and conditions of test. The assignment and training of human factors test and evaluation personnel is also considered.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Evaluation						
Performance Measurement						
System Analysis						
Field Test						
Task Analysis						

Unclassified
Security Classification

HUMAN FACTORS IN FIELD TESTING

FINAL REPORT

LS-FR-71-1

W. G. MATHENY
G. W. PATTERSON, JR.
G. I. EVANS

PREPARED FOR

ENGINEERING PSYCHOLOGY PROGRAMS
PSYCHOLOGICAL SCIENCES DIVISION
OFFICE OF NAVAL RESEARCH

CONTRACT N00014-67-C-0315

WORK UNIT NO. NR 196-067

DECEMBER 1971

LIFE SCIENCES, INC.
227 LOOP 820 NORTHEAST
HURST, TEXAS 76053

ABSTRACT

The primary aim of this study has been the development and test of methods for evaluating operator performance during field test of complex man-machine systems. Recommendations are presented which are aimed at strengthening the human factors evaluation process in both the field evaluation and the development phases of a system. However, during the study field evaluation took precedence with respect to development and trial of test procedures and techniques.

Recommendations are made based upon the study of the development and field test of specific Naval aircraft systems. Field evaluation techniques and methods were tested during the evaluation of one of these systems - the P-3C anti-submarine warfare system. These recommendations deal with the identification of measurement points, measurement scales, evaluation criteria and levels and conditions of test. The assignment and training of human factors test and evaluation personnel is also considered.

FOREWORD

This report was sponsored by the Engineering Psychology Programs office of the Office of Naval Research under Contract No. N00014-67-C-0315. It recommends methods for carrying out Human Factors field evaluation with emphasis upon operator performance within the context of the mission of the system. Such a performance oriented evaluation requires a test of the operator station design to determine whether the operator can, in fact, carry out his tasks to the criterion level required by the mission.

The authors are grateful to all those individuals and organizations who cooperated by supplying information through comments, discussion, and review of material to make this report possible. Their suggestions and the sharing of their experiences have been invaluable. The authors assume responsibility for distortions and inaccuracies which appear.

The authors express especial appreciation to Dr. M. A. Tolcott and Mr. G. S. Malecki for their technical support and guidance in carrying out the project. Special thanks are expressed also to Dr. K. F. Thomson and all of the Naval officers and men who contributed their time and knowledge to the project and to Mrs. Betty Acton of Life Sciences for coping with the problems and details of getting all of the findings and ideas into proper report form.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTD)

LIST OF ILLUSTRATIONS

1.0 INTRODUCTION

1.1 OBJECTIVE AND CONDUCT OF THE STUDY

The objective of this effort has been the development of more comprehensive and reliable means for carrying out Human Factors evaluation during field testing of complex man-machine systems. Its purpose has been to provide recommendations, procedures and methods useful for meeting the requirement for test and evaluation of military systems before acceptance for production and operational use.

While major emphasis has been placed upon field test and evaluation just prior to final acceptance of the system, the contribution and importance of test and evaluation during the entire development cycle is considered in some detail in this report.

In carrying out the project the Research, Development, Test and Evaluation (RDT&E) process within the Navy was examined in detail. After examining a number of different types of systems undergoing test and evaluation, specific aircraft systems were selected for detailed study in the development and evaluation of techniques and methods of Human Factors field testing. These were the A-7A and the P-3C ASW system. The major portion of the work was carried out using the P-3C.

In brief, the A-7A task analyses, mock-up inspection reports and field evaluation efforts were investigated in detail after the Principal Investigator had participated in the manufacturer's indoctrination course on the system. However, the major effort in testing various methods and techniques was conducted using the P-3C system. Both an Operational Sequence Diagram (OSD) and a Mission Time Line (MTL) were developed for the Tactical Coordinator's station within the context of a standard Evaluation Mission. Using both the OSD and the MTL several groups of experienced Tactical Coordinators (TACCO's) participated in tests designed to determine the feasibility of identifying operator performance measurement points within the system.

The MTL was used to test the feasibility of obtaining estimates of TACCO workload in the early stages of work with the P-3C. Although the feasibility of the method was demonstrated, the procedure is not discussed in detail in this report since the estimate of workload from the MTL is more properly a part of the system development phase.

Tests of segments of the mission were conducted in the P-3C weapon system trainer, Naval Training Device Center device 2F87, using experienced TACCO's as operators to determine the feasibility of its use at the field test level.

The detailed study of specific aircraft systems was supplemented by an extensive review of formal and informal reports and writings relevant to the problem and by interviews with human factors personnel, administrators and evaluation project personnel. A reference list of the reports found most helpful are given in the Bibliography and Source Material section of this report.

1.2 ORGANIZATION OF THE REPORT

This report considers the importance of continuous evaluation throughout the development phase of the system but concentrates upon the problem of test and evaluation just prior to system acceptance. The role of various evaluative methods applicable throughout the development and test of the system are discussed in the early sections of the report along with a frame of reference with which to approach the problem of test and evaluation. Some important personnel and organizational considerations are then taken up before the suggestions and recommendations for field evaluation proper are discussed.

In the section dealing with field evaluation processes and recommendations the importance of setting forth a standard Evaluation Mission is stated. The use of the Operational Sequence Diagram for identification of measurement points, both for manual performance and decisions is described and is followed by discussion of means of recording and evaluation including the problem of establishing criterion performance. The several test conditions under which meaningful evaluative measures may be obtained are discussed in order that the Human Factors evaluator may consider test situations other than the actual system as sources of useful and predictive evaluative data. Finally, the subjects of workload analysis, system evaluation for operator feedback and the use by the operator of old procedures and habits are discussed.

2.0 A HUMAN FACTORS EVALUATION FRAME OF REFERENCE

2.1 TOTAL SYSTEM VS. COMPONENT EVALUATION

In considering the test and evaluation of a particular system it might be assumed that the total system is in some way evaluated with respect to its capabilities for achieving some specified mission. Under this assumption it might be concluded that the field test of the system is a simple matter of determining on a " go no-go" basis whether or not the system meets the mission criteria. If such were the case, specific concern with evaluation of the human component or any other component of the system would seem to be unnecessary when carried out in connection with the test of the system.

The assumption that human factors evaluation is a necessary and somehow distinctive part of the evaluation process requires some justification and clarification. An answer seems to be needed to the question as to why there should be concern with the minutiae of evaluating components and subsystems of a system if the evaluative decision is one of "accept" or "reject" the total system. Presumably if the total system successfully accomplishes its design mission there would be no interest in measuring the performance of any of its components - human or hardware. If the system meets its mission criteria it may be assumed that the hardware and human components are functioning so as to bring about total mission success.

There are, in fact, practical and cogent reasons why component and subsystem evaluation is necessary. The reasons for conducting component and subsystem evaluation rather than an overall system evaluation, when examined closely, tend to bring into focus the requirements for carrying out an adequate human component field evaluation. The following paragraphs are intended to clarify the problem and form a rationale for the recommendations given later.

2.2 THE NATURE OF THE DEVELOPMENT PROCESS

The first important reason for being concerned with component testing during system evaluation stems from the nature of the development process itself. In theory, at the beginning of the development cycle the mission of the system is delineated in detail with criteria for successful performance clearly spelled out. Many considerations mitigate against such a clear delineation.

When a system is developed for a new mission or to extend the capabilities for executing a present mission, the details of the system and its performance criteria cannot be stated definitively at the outset. Generally, rather explicit overall system criteria

are established to be attained through application of the present state of the equipment art or the projected state of that art. However, details of the mission and the performance criteria develop with the development of the system or more precisely, with the development of the hardware for the system. As the iterative process of design proceeds the details of system components and their requisite individual performance criteria emerge. The overall goal or mission of the system is broken down into intermediate or secondary goals for subsystems and components of the total system. The attainment of these intermediate or secondary goals by the components and subsystems are intended to cascade summating in attainment of the overall system goal.

During the development process a system is being synthesized from components chosen after an analytical exercise in which the total system requirements have been broken down into subsystem and component requirements. Actually the processes of analysis and synthesis go on in iterative fashion throughout the development period. The central point, however, is that system synthesis is attained through selection of components and subsystems which can perform in accordance with the requirements made explicit by the analysis. Components are chosen (1) whose input-output characteristics match adjacent components, (2) which perform the proper transformations on the input and (3) which perform their proper function within the required time. Component and subsystem performance summate to total system performance. One needs only to reflect on the process of synthesizing a simple audio circuit to understand how the incompatibility of one component can lead to total system failure.

Field evaluation of a system takes its cue from the development philosophy. For most systems total system effectiveness in field evaluation is an estimate based upon evaluations of components and subsystems of the system. Most field evaluations, therefore, are not, and probably cannot be for systems of any complexity, evaluations of the system in toto in its intended operational environment. Rather they are evaluations of particular components and subsystems as they operate, in combination with other parts of the system in an environment more or less representative of the intended operational environment. The tested components (both hardware and human) are intended to be representative of those which will finally comprise the total system. These intentions are often only approximately realized. In field evaluation there is the need to choose the proper components and/or subsystems for test and the representative environmental conditions under which to test them if the testing is to provide an accurate estimate of how well the total system will function in its operating environment.

Since system development is a process of synthesizing components and evaluation is the testing of those components emphasis must be given to the point made by other writers and which will be discussed further in this report. That is, in order to properly evaluate a component of the total system it is necessary that the evaluator know explicitly the role of that component in the total system. The human operator is such a component. The Human Factors evaluator must know what the operator must do, how well he must do it and under what environmental conditions he must perform. This determination cannot be left to last minute speculations by the evaluator in the field.

2.3 THE TROUBLE-SHOOTING ASPECT OF EVALUATION

A second reason and necessity for evaluation at the subsystem and component level comes about when a particular chain of components or subsystems is tested and the performance fails to meet the standard. Under these circumstances it is necessary to determine which subsystem(s) or component(s) failed to perform to their particular criteria. The Human Factors evaluator is interested in determining whether the human component failed and, if so, in what way.

These circumstances require that information on performance of subsystems or components be obtained in order to diagnose the source of difficulty. This information must be obtained through a systematic and reliable means suitable for identifying the trouble spot within the larger unit <u>after</u> the larger unit has failed.

2.4 THE ADAPTIVE HUMAN COMPONENT

A third reason for component evaluation is peculiar only to the human component. It is not unusual that the ingenuity and adaptability of the human operator enables him to perform in a way which results in system success even though his actions and performance may have been quite different from those anticipated by the designer. It is this adaptability and ability to recover which characterizes the human component of the system, makes human factors evaluation (or data collection) important in all systems, and which differentiates the human component evaluation from that of other components of the system. A method for determining when and how the operator has performed in this adaptive way is necessary for guiding redesign, procedural changes or training.

3.0 LEVELS OF EVALUATION

3.1 THE ROLE AND DEFINITION OF FIELD EVALUATION

In working with the problem of Human Factors evaluation restricting the study to "field" evaluation creates a certain dilemma since it is difficult to define where such evaluation begins or ends. Although there are formally defined evaluation phases such as the Navy Board of Inspection and Survey Trials (BIS), evaluation is a continuous process beginning in the early stages of system development. Thus the field evaluation should grow out of and be dependent upon a body of information and testing of the system during its development life. It may extend well into the operational use of the system.

Further, it is necessary to define what the field evaluation is to accomplish. Broadly speaking it is meant to test the system against assumptions made about its performance as it was originally conceived. In a very real sense any test which predicts how well the system will perform is desirable at whatever point in the development process it is conducted. The element which is added to test and evaluation through conduct in the field is presumably that variables and conditions are more nearly representative of those in fleet operations and therefore more valid. This representativeness varies from system to system depending upon such factors as similarity of the system to previous systems, urgency of system need and availability of test personnel and equipment.

Tests conducted toward the end of the development cycle generally will have greater "content" validity than those conducted early. That is to say that the equipment to be tested and the conditions of test will be judged by competent evaluators to be good likenesses of the ultimate criteria, i.e., operational use. However, deficiencies found in the system at this stage are generally more costly and difficult to correct than if found earlier in the development cycle. It is desirable, therefore, that valid evaluation of component performance, human and hardware, be carried out as early in the development process as possible. The more valid the component testing early in development the fewer problems that will arise during field testing or in subsequent operational use. The most desirable situation is one in which tests carried out at all stages of development and field test have high <u>predictive</u> validity for predicting performance in operations. This <u>predictive</u> validity in which empirical relationships between tests and operational performance are established cannot be obtained until reliable measures of performance in operations are possible - something toward which more effort should be directed.

At some point before actual operational deployment an evaluation must be made to determine whether the system performs to the original conception and specifications. This is the role generally assigned

to field test. The results of this evaluation can and should also serve as criteria against which to validate tests conducted earlier in the development process. Valid testing conducted earlier in development will prevent design deficiencies from reaching the accept-reject field test point and their necessary expensive correction.

3.2 VALIDATION AND USE OF EARLIER TESTING

It has been indicated that field testing must serve to test the assumptions and specifications made during system conception and development and as criteria against which to validate testing carried out earlier in the development phase. These earlier tests take a variety of forms.

3.2.1 Physical Models

Physical models of the system range from a simple static mock-up of a part of the system through a dynamic simulator to the aircraft itself. This section discusses these models and their role in human factors evaluation.

3.2.1.1 Static Mock-Up. At present the physical model most used in human factors design and evaluation is the operator station mock-up. This mock-up is used almost continuously as an evaluative tool during the design process. It is also the focal point during a formal human factors evaluation, i.e., the mock-up inspection. It is believed that major improvements can be made in the procedures and techniques used during these inspections as they are now constituted. This belief has been found to be almost universally supported by those having experience with the mock-up inspection.

As a background for discussing the mock-up inspection and the role it plays in overall evaluation it is necessary to discuss two levels of evaluation. The first level is termed "operator station oriented" evaluation. The second is termed "mission-oriented".

In an operator station oriented evaluation the mock-up is evaluated for compatibility with the operator's capabilities and limitations. The operator station is examined for its conformity with good human engineering principles and handbook data. Evaluation is made on the basis of work-place layout, control coding, control-display relationships, illumination, anthropometric compatibility and the like.

To carry out a station oriented human factors evaluation personnel must be knowledgeable with respect to the human factors literature and data, be able to critically evaluate those data, and extrapolate from them in the light of the particular system being evaluated. Further, they must be cognizant of the principles of good human factors design as applied to the specific system being evaluated. The points to be covered during the evaluation should be incorporated into a checklist

to insure that important areas are not overlooked. The Human Factors evaluator should be knowledgeable with respect to the personal equipment to be worn by the operator and evaluations should be conducted taking into account the effect of personal equipment upon performance.

Mission oriented evaluation requires an examination of the operator station in the light of operator functions and tasks and an assessment of whether the system will perform to its designed level. In order to carry out such an evaluation it is necessary to know in detail what functions and tasks the operator is required to perform. A determination must be made as to whether the operator can carry out his functions and tasks in the proper sequence to the required accuracy within the required time. Therefore, the mission oriented evaluation requires that the evaluator have detailed information, by mission segment, about subsystem functions, data flow, and the requirements placed upon the operator. This information may be obtained through the detailed Mission Time Line (MTL) or Operational Sequence Diagram (OSD) of the system. These are discussed in more detail in Section 5.0 in connection with the problem of establishing performance measurement points.

3.2.1.2 Simulators - The simulator, as a dynamic physical model of the crew station, has several advantages over the static mock-up. From the evaluator's point of view it offers flexibility, opportunity to obtain reliable performance data and a test situation more representative of the real system. It allows for the evaluation of the dynamic man-machine interactive performance.

Advances in simulator design and construction are steadily increasing its utility in system design and evaluation. When such a device can be made available it is much to be preferred over the static mock-up for use during Mock-Up Inspection. The use of the weapon system simulator to the greatest extent possible for performance testing throughout the development cycle and during field evaluation is to be recommended. More is said about its use during field testing in Section 5.7 of this report.

3.2.2 Part Task Testing

Active concern with human factors evaluation from the beginning of the system allows for the introduction of a number of evaluative techniques. Particularly useful is the technique which has been termed the "open-loop" test of display design configurations and which may be used in weeding out or narrowing down design alternatives in the early stages of development. These techniques are essentially tachistoscopic presentations of display designs and are used for comparative evaluation. They are, in fact, the techniques through which a major portion of the human factors data available to us today was obtained.

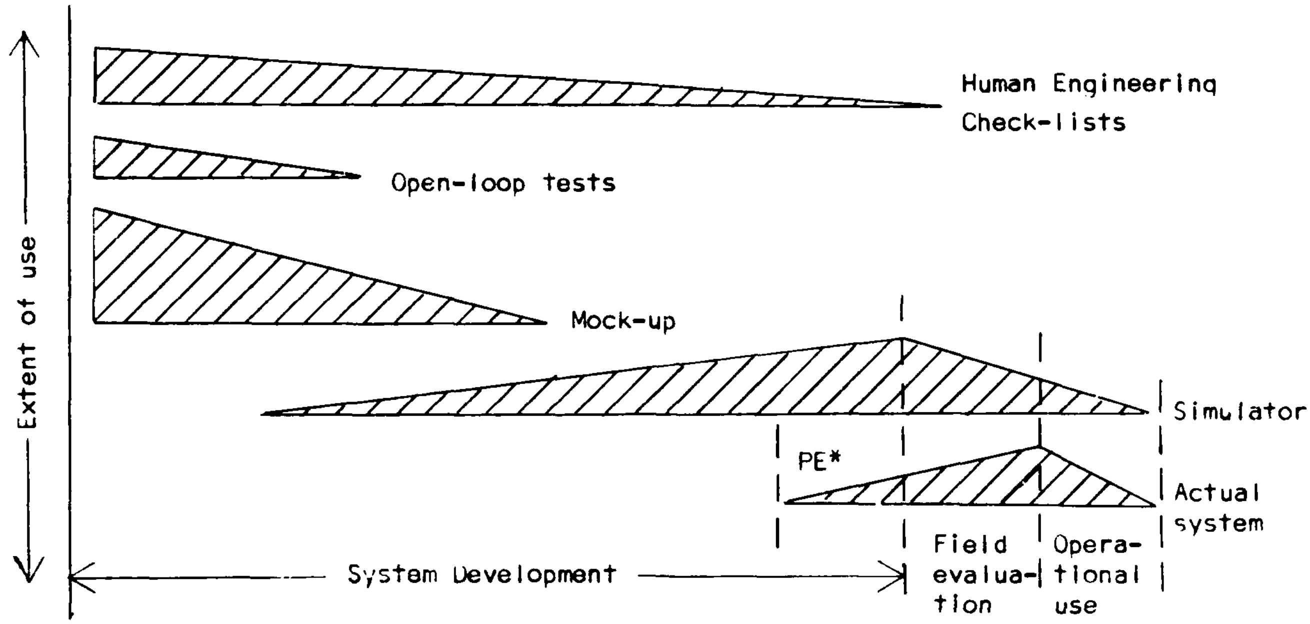
This testing technique is termed "open-loop" since the response of the subject has no direct effect upon the next stimulus presentation. The stimulus material may be a display configuration presenting information which the subject is required to interpret, readout, and report. Measures of the subjects performance may be the speed, accuracy or both with which the information is read out. The stimulus material used may be in the form of a static projection, i.e., a 35 mm slide; in such case it is termed static open-loop testing. Alternatively, the material may be presented through use of motion picture film in which display elements move realistically. This type of testing is then termed "dynamic open-loop" testing. A detailed description of this type of testing is given in Schum, Elam and Matheny (1962) in which its use has been demonstrated.

3.3 PHASING OF EVALUATION LEVELS

The initial point to be made is that human factors evaluation should begin at the point of assignment of functions in the development process and continue through the use in operations. During this evaluation process at least five methods of evaluation are applicable. These are (1) human engineering check-lists, (2) static and dynamic open-loop tests, (3) mock-up evaluations, (4) tests in the simulator of the system and (5) tests in the actual system. A suggested relationship of these types of testing throughout the development and employment of a system is given in Figure 1.

Human engineering check-lists are particularly appropriate to early stages of human engineering evaluation and can be used with the mock-up for type one evaluations. A listing of check-lists felt to be representative of those in use is given in the Bibliography and Source Material. Open-loop testing lends itself to comparative evaluations of information display components and subsystems. The simulator and aircraft are most effective in evaluating whether or not human performance meets specific criteria for system effectiveness.

Under this conceptualization the mock-up inspection as such takes on a different meaning. Under it the mock-up of the operator station as a design and evaluation tool is used on a continuing basis by manufacturer personnel under continuing monitorship of user personnel. The signifiance of and necessity for formal mock-up inspections are considerably reduced.



1. Human engineering check-lists used to evaluate design compatibility with human capabilities and limitations.

2. Open-loop tests used for comparative evaluations of display and configuration design.

3. Mock-up used in conjunction with human engineering check-lists to determine design compatibility with human capabilities and limitations.

4. Simulator used to evaluate operator performance against performance criteria. In operational use it may be used in:

   - Evaluating proposed new or modified tactics
   - Accident investigation
   - Diagnosis of performance which is below criterion requirements

5. Actual system used to evaluate against performance criteria and to evaluate proposed new or modified tactics.

* PE - Preliminary Evaluation

Figure 1. Relationship and extent of use of test procedures during system development and evaluation.

4.0 PERSONNEL AND ORGANIZATIONAL CONSIDERATIONS

4.1 DOCUMENTATION

In the recent past the requirements for consideration of human factors in the design and evaluation of military systems has become more explicitly documented. This comes about through adoption of MIL-H-46855, Human Engineering Requirements for Military Systems, Equipment and Facilities, 16 February 1968 and MIL-STD-1472, Human Engineering Design Criteria for Military Systems, Equipment and Facilities, 9 February 1968. These two documents when cited in a contract specification provide authority for carrying out effective human factors effort. However, the documents themselves cannot spell out the requirements in the detail which ensures accomplishment of a good human factors effort in the absence of trained personnel in sufficient numbers dedicated to making their contribution felt.

4.2 MANAGEMENT

For human factors evaluation, assignment of the right personnel begins at the project office level with responsibility for the evaluation function vested in a designated individual. It should be his responsibility to see that the system and time line analyses contain the information on test points and their priorities (see Section 5.0). These test points should be within the context of an evaluation mission. This mission, set forth early in development, will evolve in detail as the equipment configuration becomes firmed up. It is the framework upon which both Human Engineering design and evaluation will hang. The human factors engineer in the Project Office must insure that this mission and the evaluation test points are developed.

4.3 ASSIGNMENT OF CUSTOMER PERSONNEL TO CONTRACTOR FACILITIES

The assignment of customer personnel to the human factors effort at the contractor's facility is highly recommended. It is suggested that, optimally, these personnel be graduates of specialist schools with special additional training in human factors and recent experience in systems similar to the one being evaluated. The importance of the human component to the system and the scarcity of hard data on man's performance in such systems warrants giving especial attention to the qualifications of the design and evaluation personnel who deal with him.

These customer personnel should be selected for their interest in and any special qualifications for human factors work. Analogous to the practice of test pilots having the additional qualifications of aeronautical or other engineering degrees the human factors

specialty should require special qualifications in human engineering. A special course of three to four weeks minimum should be made available for this specialty coming after the individual has qualified as a specialist, e.g., a Navy Test Pilot. This special training course is discussed further in Section 4.5 and a suggested course outline is given in Appendix A. This course is suggested as a minimum requisite and not a substitute for formal training in Human Factors. The responsibilities of these personnel at the contractor's facility would be purely advisory in helping to develop the evaluation mission, critical test points and methods of measurement. They would become members of the evaluation team during mock-up inspection lending their detailed knowledge of the operator requirements and the data flow through the system to a more objective and mission oriented evaluation.

4.4 CONTRACTOR ASSIST DURING CUSTOMER EVALUATIONS

During evaluations carried out at the customer's facilities it is recommended that contractor human factors personnel assist and advise on-site in the planning and conduct of the tests. These contractor personnel would contribute their knowledge of the system to the planning and conduct of the tests. Their detailed knowledge will help the customer evaluator immeasureably in working out the details of how to determine test points and in diagnosing sources of operator difficulties.

4.5 TRAINING OF EVALUATION PERSONNEL

The approach to be taken by the evaluator in any evaluation is essentially that which the serious experimenter would take in testing an hypothesis. He must be as knowledgeable and have as much quantitative information as possible about the variables and conditions influencing the operator behavior. He must either control these or be able to assess their effects. He must also have an understanding of experimental design, of reliability of measurement and of data analysis and report. An appreciation of these requirements coupled with experience and knowledge of the operational conditions under which the system will function would combine to maximize the effectiveness of the human factors evaluator. Their combination in a single individual is rare. A training program designed to produce such a combination is recommended.

The human factors engineer usually comes to the evaluation situation with a limited knowledge and appreciation of the operational demands. The customer project personnel assigned to evaluate the system have the operational experience but usually are not experienced in the methods of experimentation which should be applied. A cross training program is recommended.

It us suggested that customer project personnel could become oriented and minimally knowledgeable about experimental methods through an indoctrination course of three to four weeks minimum. This course would be offered to project officers and project pilots who are directly concerned with planning and conducting the evaluation. The course outlined in Appendix A is suggested. After such a course

a person could not be considered an authority in all areas of evaluation. Rather, he would gain an overall perspective of the requisites for evaluation and appreciate the need for consultation and assistance from subject matter specialists in such areas as experimental design, performance measurement and rating, and analysis of data.

Human factors personnel should be indoctrinated in the operational use of like systems in every way possible. For example, within the Navy it has been suggested that short tours aboard carriers by human factors personnel should be undertaken. Every opportunity for these personnel to observe the operation of similar systems either in the operational theater or in training operations should be taken.

5.0 FIELD EVALUATION PROCESS AND RECOMMENDATIONS

5.1 OUTLINE OF THE FIELD EVALUATION PROCESS

As indicated in earlier sections of this report the field evaluation is mission oriented and is designed to assess whether the system will perform to some specified criterion level. In order to carry out such an evaluation for the human component it is necessary to know in detail what tasks the operator must perform, in what sequence, and to what criteria. It is also necessary that the operating environment of the system be thoroughly understood so that the important aspects of that environment may be incorporated into the test and evaluation process. A determination must be made as to whether the operator can carry out his functions and tasks in the proper sequence to the required accuracy within the required time under conditions representative of the operational environment.

The suggested steps for carrying out the Human Factors field evaluation are given in Figure 2. A brief discussion of the items shown in this figure is given as an orientation to the overall human factors test and evaluation process.

Normally Steps 1 through 4 and 8, 9 and 11 will have been accomplished during system development as basic information for design and evaluation and will be available as data for planning the field evaluation. If they have not it will be necessary that the best approximation to them possible be carried out early in the field evaluation phase.

The Evaluation Mission of Step 1 in Figure 2 is discussed in detail in Section 5.3. This mission is oriented toward the tactical use of the system and must incorporate those design elements involved in its tactical employment.

The sequential task listing of Step 2 is derived by detailing the actions of the operator within the segments of the evaluation mission. This step is detailed in Section 5.4.

In Step 3 the sequential task list is placed in the Operational Sequence Diagram (OSD) format. This is described in Section 5.4. The OSD will be of primary use in determining whether tasks have been omitted, whether they are in the proper sequence, and is recommended to be used in Step 10 in identifying operator performance measurement points.

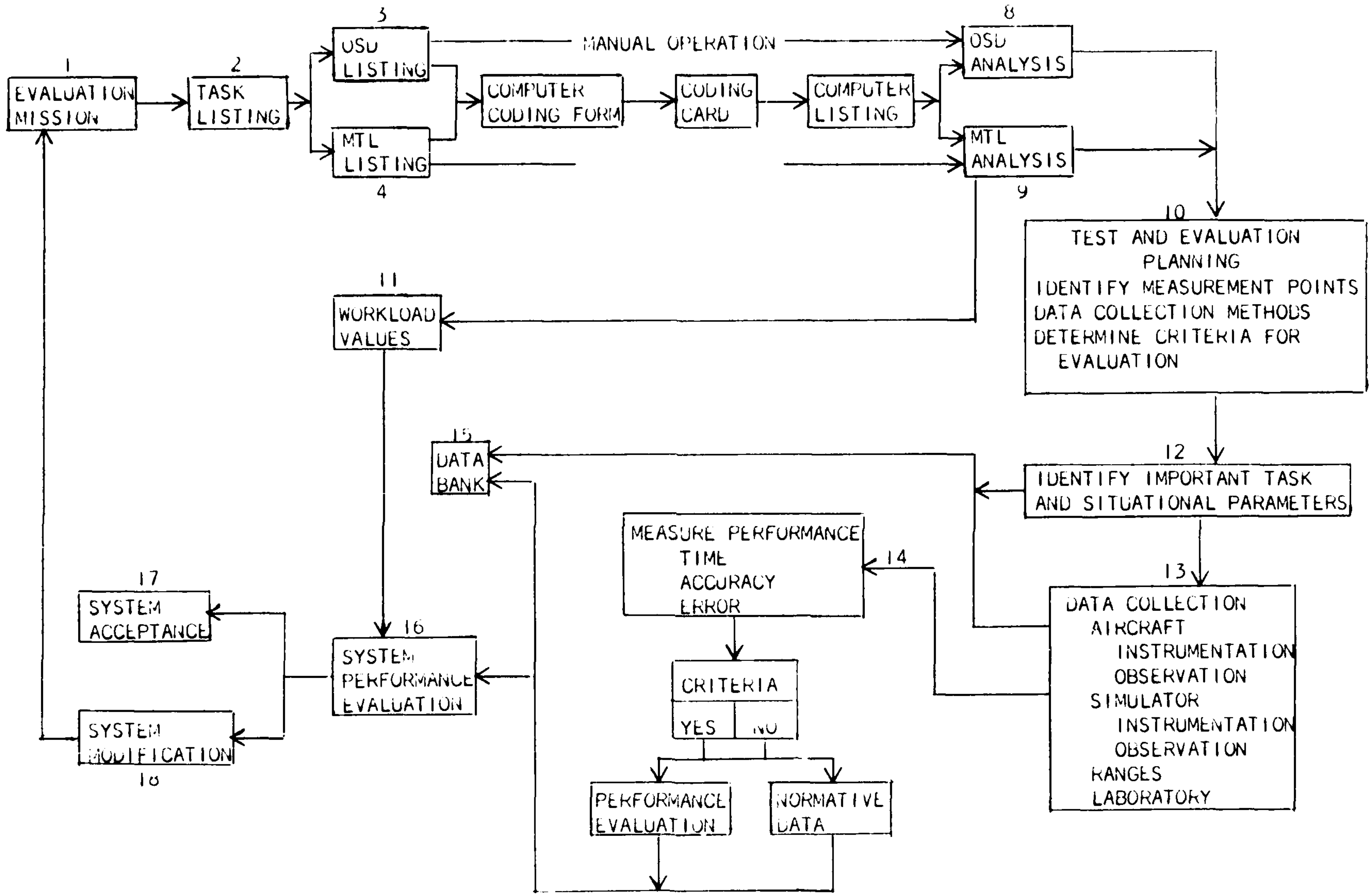


Figure 2. Recommended procedure for measuring and evaluating operator performance

In Step 4 the task sequence derived from the OSD is transferred to a Mission Time Line (MTL) format. The MTL provides a check on the proper sequencing of operator tasks and provides the basic information for deriving the workload values of Step 11.

When computer equipment and facilities are available, Steps 5, 6, and 7 are highly recommended. This process is discussed in Section 5.4. Computerizing the generation of the OSD and the MTL enables the evaluator to incorporate most quickly the changes that occur in the system equipment and task sequence during evaluation.

In Step 10 of Figure 2 the test and evaluation procedures for measuring operator performance are detailed.

The *Operational Sequence Diagram* is used for identifying points within the system at which the operator and his equipment interact and at which performance may be measured. The characteristics of each of the points selected will govern the type of data to be collected and will provide insight and information as to how the data may be recorded and analyzed. This process is covered in Section 5.6. Methods of collecting and recording data at the measurement points which have been identified are also discussed in that Section.

The problem of determining the evaluation criteria for each of the measurement points selected is discussed in Section 5.6.3. The measures taken in Step 14 are evaluated using the specified criteria for successful performance. The steps to be taken when no criteria are available are also discussed. This evaluation leads to either system acceptance, Step 17, or system modifification, Step 18. Where system modifification occurs the changes in the evaluation mission and the operator's tasks are incorporated into the OSD and MTL and the appropriate re-evaluation of the operator performance made.

5.2 SOURCES OF INFORMATION

In order to evaluate the performance of the human operator during field evaluation of a system, it is important that certain information be available describing the requirements placed on the operator by the specific equipment being evaluated. The major portion of this information normally would be generated during weapon system development since this process requires a knowledge of or projections about (1) mission requirements, (2) equipment capability, (3) vehicle performance and limitations, and (4) crew capabilities and limitations. However, continuous updating of the information is necessary through the test and evaluation phase as the components or subsystems of the system may be changed, equipment may not perform to expectations, and tactics and procedures may be modified.

The Human Factors evaluator must be well informed about the system requirements and system design. Only through a thorough knowledge of the system can he create or update the task analyses. His work will be affected continuously by changes in system design and equipment capability. He must be equipped to incorporate these changes into his analysis readily and modify his test plan accordingly. If the evaluator has worked closely with the design and test of the system through its development, he will be well prepared for the field test phase in respect to knowledge of the system. Normally he will not have this experience.

The methods by which information about the system can be obtained by the evaluator are (1) study of relevant documents, (2) interviews with contractor design personnel and (3) interviewing of experienced operators. A background of experience with previous versions of the system or with earlier systems for accomplishing the mission will be extremely beneficial in understanding and evaluating the system. The relevant documents include but are not limited to the system and component specifications, basic mission manuals and contractor publications. System design information may be obtained from contractor documents and through interviews with contractor personnel. A knowledge of the operational environment of the system may be obtained from the requirements documents, through interviews with operators experienced in similar systems and from contractor publications reciting and assumptions made during design about the operational tactics and environment.

5.3 THE EVALUATION MISSION

The purpose of the evaluation mission is to provide a standard test situation which incorporates as many of the design elements of the system as possible and which represents the operational system with respect to tactics and environment. For systems in which several unique missions may be accomplished multiple evaluation missions will be necessary. If a detailed design mission has been developed during the system development phase the evaluation mission should parallel it closely. The evaluation mission may need to be varied from the design mission because of available test facilities and time and budget limitations.

For most systems being evaluated emphasis should be placed upon the tactical phase of the mission since it is in this phase that goals can be pin-pointed and the success or failure of the system is determined. For example, in the F-3C system the tasks involved in taxi, take-off, climb-out, cruise, let-down and landing are necessary to accomplishment of the mission. However, for this system emphasis would first be placed upon evaluation during the tactical phases of the mission in which the ASW equipment is being exercised.

Generally, the evaluation mission profile will begin with entry into the tactical phase. All of the factors that affect the successful accomplishment of the mission must be considered. If possible the evaluator should discuss the evaluation mission design with operator personnel experienced in missions similar to the one in question to insure that it is feasible and representative of the operational requirement. The tactical phase of the mission should be subdivided into segments and may be further subdivided into events. These division points may be somewhat arbitrary but are generally determined as being at the point of completion of a definite series or sequence of actions for which a definite beginning and ending time can be established.

With most complex systems there are times when some part of the system is not functioning properly. However, in order to evaluate the system as it was designed to perform and to provide a reference for comparing degraded performance to design performance, all components of the system must be assumed to be operating properly in the creation of the evaluation mission and the operator task sequence. From this base the effect of various equipment malfunctions upon the operator's task and mission performance may be evaluated.

5.4 DERIVATION OF OPERATOR TASK LIST AND CREATION OF THE OPERATIONAL SEQUENCE DIAGRAM (OSD)

An outline of the mission profile must be completed prior to starting the operator task listing. The task list is developed within the framework of the mission and is a step-by-step description of the operator's tasks as the mission proceeds. If the evaluator has available a task listing for the design mission he must first modify the design mission, where necessary, to a specific evaluation mission which is feasible and representative. He must then modify the task listing according to this mission. When no detailed task listing for a design mission is available the evaluator will find it necessary to develop both the evaluation mission and the task listing "on the spot" in order to understand the system in enough detail to decide upon operator performance measurement points. Whether modifying a design mission or creating his own the operator tasks of interest to the evaluator will be determined by equipment characteristics and mission requirements. The task listing for the evaluation mission should include only those which are mission essential; that is, those tasks that must be completed successfully in order to achieve the prime mission objective.

The Operational Sequence Diagram (OSD) format, Appendix B, lists the operator tasks and diagrams the interaction of the operator with his equipment and with other operators. Again, for the purpose of identifying points at which to measure operator performance, the OSD should contain only those tasks essential for executing the mission. It should show when the operator moves

a control or actuates a switch, the equipment response, the operator's observation of the response and his further action. In some cases the operator's response will be to verbal instructions or information from other operators. His action may be a manual action or a verbal directive.

The OSD shows graphically the branching and feedback relationships between tasks and the interactions of the operator with other components. The illustrations of the OSD format given in Appendix B and in MIL-H-46855 of 16 Feb 1968 are useful for learning the mechanics of the OSD. The OSD shown in MIL-H-46855 uses geometric symbols to code actions and behaviors of the operator and his equipment while the example in Appendix B uses letters for this code. These letter codes are also given in Appendix B.

The present writers found the letter code to be somewhat easier to use than the geometric symbols in the generation of the OSD. The letter coding allows also for the use of a computer generated task listing as described by Wilson (1968). The reports by Wilson (1966, 1967 and 1968) are highly recommended reading for becoming familiar with the Operational Sequence Diagramming procedure. The obvious advantage of a computer generated OSD is the relative speed with which it can be changed when changes are made in the equipment or operating procedures.

In creating an OSD it should be borne in mind that its fundamental purpose is to depict the interaction of the operator with other components of the system. In this interaction the operator makes an input to the other system components - either equipment or other operators. These components react in turn and information is conveyed to the operator concerning the results of his inputs. The coding and sequential diagramming of these interactions is the function of the OSD.

5.5 USE OF THE OPERATIONAL SEQUENCE DIAGRAM IN THE IDENTIFICATION OF MEASUREMENT POINTS

Five basic steps are considered necessary for the measurement and evaluation of human operator performance in the system during the field evaluation.

1. The identification of manual performance measurement points, i.e., the identification of those points in which the operator interacts overtly with other components of the system and at which quantitative measurement of his actions can be made.

2. Identification of tactical decision making points within the operation of the system.

3. Determination of the type of data that can be collected and the appropriate methods of recording at these measurement points.

4. Determination of the performance criteria, i.e., the level of performance required of the operator at these measurement points and how it may be stated.

5. The evaluation of the operator's performance, i.e., the comparison of his measured performance with the design criteria resulting in a judgment as to the adequacy of operator performance.

The OSD is used as a primary aid in the identification of the measurement points within the system (Steps 1 & 2). This may be done by a Human Factors evaluator familiar with the system or by an operator who has been trained for participation in the field evaluation. If the Human Factors evaluator has worked with the system during its development he would normally be involved in either OSD or MTL development, be familiar with the system and have already isolated a number of the measurement points. If the evaluator has little experience with the system he may ask the trained operators to identify the measurement points through studying the OSD.

The objective in selecting manual performance measurement points, is to identify those points at which there is a culmination of activities of the components and at which the effect of their cumulative performance over the preceding sequence can be measured. This may be conceived as being a series of related activities funneling down to a point at which the operator acts upon the information received. These are accomplishment points at which there is an overt and observable interaction between the operator and the machine or between operator and operator. They are points at which information flows between components and at which operator errors may occur.

The OSD provides the basic data for identifying manual performance measurement points. It is recommended to the individual picking these points as the means for providing him with a detailed picture of the tasks carried out by the operator. The evaluator studies the OSD and from his knowledge and experience with the system selects those points at which the operator overtly and observably makes an input or action in operating the system.

The OSD also provides the basis for identifying those points in the operation of the system at which decisions important to the success of the tactical mission are made. These decisions may or may not result in a direct overt action - verbal or manual.

Although the points at which these decisions are made may be identified from the OSD in a manner similar to the identification of manual performance measurement points, the determination of their correctness, i.e., the establishing of a criterion of adequacy of the decision, is generally more difficult than is the case for performance measurement. This problem is discussed in more detail in Section 5.6.

In carrying out the studies from which the recommendations in this report are drawn both the Operational Sequence Diagram and the Mission Time Line methods of analysis were used on a trial basis for the identification of measurement points. In these trials experienced operators of the system were asked to study the analyses and to indicate performance measurement points. It was found that these operators, being oriented toward performing tactical operations, tended to select tactical decision points rather than actual interactive manual performance measurement points. They tended not to zero in on the more minute performance interactions between the operator and the equipment at which errors could be observed directly and measurements of performance taken. Rather, they tended to emphasize tactical decision making activities. Special and explicit instructions were necessary to orient them toward the more minute manual performance measurement points.

The results of present work with the OSD in identifying measurement points underlines the need to emphasize the point that experienced operators will tend to focus on tactical decision points and be less inclined to pick more minute man-equipment interface performance points. It is important that both types of measurement points be identified. Proper design of the points of interface will ensure that errors in information flow across the interface do not occur. However, it is also necessary to determine whether the decisions based upon this information can be made correctly. Therefore, both manual performance measurement points (points of man-machine interface) and tactical decision points need to be identified. It is recommended that where possible, Human Factors engineers familiar with the system identify the manual performance measurement points and the identification of tactical decision points be made by operators experienced with the given or similar system. If the experienced operator must be called upon to make both types of identifications, it is recommended that he first go through the OSD with the sole purpose of identifying manual performance measurement points. He should then go through the OSD a second time with the purpose of identifying tactical decision points. Again, instructions emphasizing the distinction between the two types of measures are especially important when experienced operators of the system perform the measurement point identification exercise. An illustration of the format, instructions and results for such an exercise is given in Appendix C.

In making his identification of performance points the evaluator or operator should avoid making judgments as to whether the points he selects are critical to system operation. He should approach the task of identifying points with the orientation that every action in a system is necessary to the success of the system. When all measurement points have been identified they may be rated with respect to relative criticality if it is found that it is not practical to record or observe at all points identified. However, where possible performance records should be taken at all measurement points which have been identified.

5.6 PERFORMANCE RECORDING AND EVALUATION

After identifying the points in the operation of the system at which performance measurements may be taken, the Human Factors evaluator must decide what measurements to take, in what testing situation to take them and how to evaluate the measures once they are collected.

In determining what measurements to take the HFE must decide both the kind of behavior to be recorded and the particular method or technique for recording it. He must also decide upon a scale by which he will quantify the behavior and he must determine the criterion or level of performance required. In this section four task types are described and scales of measurement, criteria, recording techniques, the evaluative judgment appropriate to each, and test conditions are discussed.

5.6.1 Task Types

The evaluator may distinguish three different types of manual task performance which require somewhat different measurement scales, criteria and methods of recording. These are (1) continuous control activities in which the operator tracks a moving target or nulls an error through continuous control movement, (2) discrete, control activities such as pressing a button or throwing a switch and (3) positioning a pointer or marker with reference to some index such as setting a dial or positioning a symbol on a scope face.

The fourth task type which the HFE must consider is that of tactical decision making already mentioned in Section 5.5. These decisions may be of the obvious simple yes-no-variety. For those decisions which appear to be highly complex or appear to require choosing among a number of alternatives the evaluator should analyze the sequence of events to determine whether the seemingly complex set of options cannot be reduced to a series of simple "yes-no" decisions. Often they can.

5.6.2 Measurement Scales

In selecting measurement scales, reference to the detailed description of the tasks of the Operational Sequence Diagram will aid the evaluator in making decisions about what type of scale is appropriate for each task. Table I, Column 2 summarizes briefly the scales appropriate for the different task types.

5.6.3 The Criteria

With the identification of measurement points within the system, it is necessary to determine whether criteria are available for these points. These criteria must set forth the tolerances or limits of required performance in the scale of measurement appropriate to the performance.

The determination of the criterion level of performance at the selected manual performance measurement points will usually require choosing among several different approaches. Theoretically (and ideally) the system designers specify the accuracy with which the human operator is required to perform. Usually, this ideal is only partially attained.

At those measurement points at which categorical binary decisions or discrete manual actions are required the criterion conditions can usually be stated directly, i.e., the operator is expected to exercise the correct option each time. At those points at which a distribution of errors is possible neither the allowable limit of error nor the expected error distribution may be stated explicitly.

The Human Factors evaluator may obtain useful information as to allowable error limits by questioning the system design engineers and/or through his own further detailed analysis of the system. For each identifiable manual performance measurement point he must ask "what are the limits of accuracy within which this task must be performed in order for the mission to succeed?" For some measurement points he will not be successful in defining the limits. For these the design engineer may have made a judgment, based upon his experience in designing similar systems, that the operator would be able to perform adequately and that the effectiveness of the system would increase with operator proficiency.

When no explicit criteria are available for a given measurement point it is necessary to record the performance of the operator at that point in "raw data" form without reference to criteria. These raw data provide the basis for estimating the effect of operator performance at that point upon total system performance. An estimate of the reliability of the operator in performing at that point is obtained which can be used in computing total system reliability estimates.

TABLE I.

SUMMARY OF TASK TYPE, SCALE OF MEASUREMENT, CRITERIA AND EVALUATIVE YARDSTICK

Task Type	Scale of Measurement	Criteria	Evaluative Yardstick
Continuous control	United appropriate to description of the task, e.g., degrees, knots, feet, yards or miles.	Established design limits within which dimension must be controlled.	Percent of time out of design limits
		Standard which must be held	Average absolute deviation from standard
Discrete control movement	Categories	Correct position established by design	Percent of responses incorrect
Positioning of pointer or symbol with reference to an index	Units of the scale in question, e.g., degrees, units, tens, hundreds	Established design position or design limits about position	Percent settings not on design setting or percent outside design limits
Tactical decision	Either-or	Judgment of panel of experts as to correct decision	Percent agreements with judges established correct response

In collecting "raw data" as well as collecting data relative to a criterion it is necessary that the performance be measured repeatedly under a given set of conditions so that a reasonably reliable data base may be obtained. The goal is to obtain as accurate an estimate of the real distribution of errors as possible through sampling the performance of a number of operators. Of course, the larger the number of operators sampled on a given task under a given set of conditions the better the estimate (or prediction) of what the performance will be when the system is operated in the future. In carrying out tests for purposes of predicting future performance it is important also that the conditions of test and the operators tested should be as representative as possible of those to which prediction is to be made.

In the field test situation, difficulty may be experienced in making repeated tests under the same conditions. This is particularly true of testing in the fully operational system since it is difficult to maintain standard testing conditions from test run to test run. The use of other test situations such as a simulator of the actual operational system has certain advantages for repeating standard test conditions and accumulating data over a greater number of test runs. The use of simulators as well as other test situations are discussed in Section 5.7. Whenever it is possible to measure performance reliably in the actual system and to take measures during at least 5 identical test runs, a reasonable data base for predicting future performance will have been obtained. For a review of the techniques of sampling and of estimating population parameters see Johnson (1949, pp. 104-117), and Cochran (1964, pp. 18-26).

The criterion or acceptable level of performance for tactical decisions is often difficult to define. The correctness of a decision at a given point in the system operation may not be ascertainable until a later time after subsequent events have unfolded. Events and decisions subsequent to the decision in question may make the assessment of its correctness virtually impossible. Some success may be attained in breaking down the decision process into a component series of yes-no decisions each of which can be evaluated as to its correctness. Where a playback record of the events of the test run is available, as for example in the P-3C ASW system, a panel of experts may view the playback and judge the adequacy of the decision. More is said concerning this problem in Section 5.6.7 in which the measurement and evaluation of tactical decisions is covered.

5.6.4 Measurement and Evaluation of Continuous Control Tasks

Continuous control tasks may be of two types. One is termed compensatory control in which the operator seeks to hold an indicator in alignment with some fixed index and, through moving the control, compensate for disturbances which tend to move the indicator away from the index. An example of a compensatory control task is the holding of a given aircraft bank angle using the attitude indicator. The second type of continuous control task is the pursuit task in which the operator seeks to hold an indicator in alignment with an index which is moving. An example of this type of task is the positioning of a symbol over a moving target on a radar scope through movement of a control.

In continuous control tasks the measurement scale problem is that of choosing an appropriate metric (inches, degrees, etc.) by which the amount of deviation of the indicator from the index may be quantified. These deviations are measured and recorded from moment to moment and provide a quantitative expression of task performance. Comparison of this performance with the criterion, or required level of performance, is the essence of the evaluation process.

The criterion or required level of performance may be given in terms of (1) a standard to be held such as a given altitude, or (2) limits about a standard within which the system is to be controlled, e.g., ± 50 feet of altitude. As indicated in Table I the type of measurement taken (measurement scale) will be different for the two types of criteria. When the criterion is given in terms of holding a standard the appropriate measurement scale will express, through some numerical value, the amount of deviation or error about the standard. When criterion limits have been set up about some standard, the amount or percent of time outside those limits can be recorded as a numerical expression of performance. For these measures either a distribution of errors or time out of limits for all operators over all test runs can be generated. When no time out of limits or no errors larger than the criterion limit occur, the evaluative decision is straightforward. The operator is performing to the criterion. If the error distribution contains errors which are greater than the criterion limit the frequency or percentage of these errors form the basis for predicting the magnitude and frequency expected during future operation of the system and for calculating their effect on total system performance. The disposition of the case in which performance outside of limits occurs may take a number of forms. The problem may be corrected by redesign of equipment, further training or re-training of operators, or changes in procedures in which the role of the operator is reordered.

For measuring deviations from the standard or for determining time out of limits for continuous motor tasks it is desirable to use some method for recording error continuously. In some systems traces or records of parameters of performance of the system may be obtained since they are generated and used as a part of the operation of the system. For example, in the P-3C ASW system various aircraft performance parameters are sensed and used in the solution to the ASW problem. Some form of magnetic tape recording is optimum for recording such parameters when summarization and analysis of the data can be carried out using ground based computers.

When limits criteria are given, a hard copy pen recording of performance may be adequate since a template or limit lines may be used for obtaining measurements of time out of limits with a fair degree of ease. A less desirable method of recording is a film record of the operator's instruments or displays from which the evaluator may sample and measure the error over the period of the performance. Finally, the evaluator may act as an observer and merely sample the performance on a periodic basis and record it manually. This latter method is quite satisfactory when the deviations of the system from the standard are of low frequency, i.e., error accumulates slowly, so that the observer has adequate time to sample and record the performance at periodic intervals. The photographic and observation techniques can be combined when motion film of the performance is taken and later reviewed by one or more evaluators. During this review observations may be made at pre-set periodic intervals. This combination is often the most expeditious and least expensive method for obtaining continuous manual performance data. It provides hard records of performance with the least expense and often with the least interference with the operation of the system.

5.6.5 Measurement and Evaluation of Discrete Manual Tasks

Discrete manual tasks are those in which the operator makes a discrete motion either to push a button or to position a multiple position switch. The recording problem here is one of recording an event. The criteria may be of two types. In the first the button is either pressed or the switch is positioned correctly. In the second the button or switch is activated in the correct sequential position relative to other discrete motor acts.

The recording problem with respect to discrete tasks is somewhat simpler than the problem of continuous motor performance recording. The discrete action may be recorded as an event either on magnetic tape or on a hard copy oscillographic record. The accuracy or appropriateness of the action can be determined through either computer processing of the tape record or visual inspection of the hard copy. With this type of task a film or video tape record is a useful recording technique since the evaluator can determine through observing the recording whether the actions were correct and in the proper sequence.

A particular advantage of the film or video tape record is the coverage of a number of controls through one recording medium. The opportunity to observe and evaluate the performance away from the actual working system and to re-play the action is also advantageous. The solution to the problems of positioning cameras and the technical problems of obtaining readable film are difficult in some systems. However, where possible of being used, the photographic or video techniques have much to recommend them since they capture a great deal of data for later reduction, analysis and study.

With the discrete manual tasks and the dial setting tasks discussed earlier, direct observation and manual recording by the evaluator often may be the most expeditious and satisfactory means of recording performance. When the evaluator is quite familiar with the task demands and is practiced in observing them he can cover a great many discrete and dial setting task performances quite reliably.

5.6.6 Measurement and Evaluation of Dial Setting Tasks

As with the discrete manual tasks the dial setting tasks are discrete events which can be recorded as events either through use of recording equipment or through observation by the evaluator. The criterion may be either a required dial setting or the carrying out of the event in the correct sequence with respect to other discrete events within a series of events. Recording may be accomplished either through recording on magnetic or hard copy tape or through film or video recordings. In general they may be treated as discrete motor tasks except that there may be specific requirements with respect to the accuracy with which the dial must be set. In this respect they are like the continuous tracking tasks in that the frequency of settings outside of limits must be ascertained.

5.6.7 Measurement and Evaluation of Tactical Decisions

The recording of tactical decisions is quite similar to recording discrete manual tasks when the decisions are observable as overt verbal or motor acts. The complication in evaluating such overt decisions comes from the difficulty in establishing the criteria as to their correctness. However, the tactical decision may be "not to act." In this case the fact of the decision and its degree of appropriateness may be evident only after the occurrence of a number of subsequent events.

To further complicate the evaluation of tactical decisions, they are made in a dynamic evolving situation in which the antecedent conditions to the decision are almost never the same from one test run to another. This is particularly the case when evaluating in the actual operating system. Antecedent conditions during evaluations can be controlled much more satisfactorily in the dynamic simulator of the system.

In making judgments as to the correctness of tactical decisions it should be emphasized that the correctness of the decision is a function of both the accuracy of the data received by the decision maker and the appropriateness of his interpretation and analysis of the data. For example, in the ASW system an incorrect tactical decision may come about as a result of improper assessment of the situation on the part of the Tactical Coordinator or as a result of inaccurate information supplied to him by other components of the system.

Further, the assessment of the correctness of a given decision "on-the-spot" at the time it is made is often not appropriate since the validity of the decision must be established in the light of the events following it. Knowledge of the information upon which the decision was made, the decision itself and the effect of the decision upon mission success are all necessary to judging its adequacy.

Within some systems the events surrounding the making of the tactical decision may be recorded in enough detail to make reconstruction or "play-back" of the events possible outside the operation of the system. Under these conditions expert observers may make a judgment as to the correctness of the decision. In the P-3C ASW system, for example, a number of records are taken which can be played back for use in determining the adequacy of tactical decisions. In other systems the records taken to evaluate the performance of the system equipment may be "played back" and analyzed to determine the correctness of operator decisions made during the evaluation mission.

Any knowledge of operator action, instrument information and verbal interchange can be used to advantage by expert observers to arrive at judgments as to the correctness of decisions. Film and video recordings, sound recordings and magnetic or other recordings of events and performance must be used to the extent available in evaluating the tactical decision. Examination of the Operational Sequence Diagram to identify the points of information exchange and points of decision is recommended as being essential to the determination of points of measurement, types of recording and scales of measurement.

When tactical decisions are judged to be in error it is necessary to examine the system both with respect to the adequacy of the equipment design and the complexity and sequencing of the data which the operator is required to interpret and analyze. As indicated above, two sources of variability contributing to decision error may be identified - information error and assessment error. A re-examination of the task descriptions in the Operational Sequence Diagram and the manner in which they are sequenced may reveal design inadequacies which are contributing to the tactical decision error.

The evaluator may find it necessary to set up tests in other environs such as the laboratory or simulator in order to isolate the contribution of information error to the total tactical decision making error.

5.7 TESTING CONDITIONS

What are the test conditions under which operator performance measurements may be taken? During field test and evaluation the goal is to obtain information about the effectiveness of the system under conditions representative of its functioning in operational use. From these data a forecast may be made about how the system will function in the operational environment. Therefore, the Human Factors evaluator attempts to obtain data on operator performance under conditions which will be most predictive of its performance during operations. The validity of his data will depend both upon how well his test situation represents the operational conditions and upon the reliability and discriminating power of his measurements. The degree to which the test conditions can be representative of operational conditions and the reliability of the measurements taken will vary greatly from system to system. The complexity of the system and the facilities and personnel made available for the test will be important influences on the adequacy of the tests.

Often it will be necessary to compromise between test conditions which provide reliable measures under repeatable conditions and those which provide the most representative operational conditions. In general, the greater the number of operational parameters incorporated into the test situation the more difficult it is to maintain standard test conditions from test run to test run. In like fashion, the more control exercised over the test situation in order to attain repeatable conditions the more unlike the operational conditions it may become. This is not a necessary state of affairs but generally, the more variables operating in the situation to which one wishes to predict the more difficult it is to control them so that standard test conditions may be attained across a series of test trials.

In the operation of many military systems many important variables are simply not controllable and/or manipulatable within the test situation. Variables such as wind speed, turbulence and visibility may affect some systems so that controlled test conditions are impossible. In such cases the evaluator must obtain an assessment of the level of each of these variables during the test run so that their effect upon performance may be estimated - through calculation of correlation coefficients where sufficient data are obtained or subjectively where it is not.

The Human Factors evaluator should bear in mind that <u>it will be of limited value to conduct tests in a highly realistic and representative test situation if it is not possible to obtain repeated measures of task performance under the same test conditions.</u> On the other hand reliable measures taken in a non-representative test situation are equally limited.

Keeping in mind that he wishes to maximize his predictions of actual operational performance from his test data the evaluator should consider at least four different test situations as sources of evaluation data. These are (1) the actual operating system, (2) dynamic simulators of the system, (3) test ranges, and (4) special laboratory-like assemblies of subsystems and parts, or simulators of parts, of the system, These test situations are discussed in the following sections.

5.7.1 The Operating System as a Test Situation

The test condition most representative of the operational use of the system during which evaluation data may be collected is that in which the full operating system is carrying out the standard evaluation mission. In the actual conduct of human factors evaluation during this mission the performance measurement may require that the testing be fitted into the testing of the system equipment. In many instances human factors test and evaluation can be worked out satisfactorily under these conditions. However, the scheduling of operational runs for the prime purpose of obtaining human factors data will be necessary for fully adequate conduct of such testing.

While operation of the actual system is the most representative condition during which to measure performance the problems to be encountered should be appreciated. The major deterrent to good test procedure is the difficulty in obtaining measures under repeated test conditions in which the important variables in the testing environment are kept the same from test trial to test trial. When the test conditions vary from trial to trial a reliable picture of operator performance cannot be obtained. Without reliable measures of operator performance under the test and evaluation conditions valid predictions of how he will perform in the operational situation are not possible. Further, unless such reliable measures can be obtained in actual full system operation the validity of the evaluative decisions made earlier in the development of the system cannot be determined. With attention to the requirements for reliable testing during the planning and staffing for field test and evaluation the repeatability of test conditions can be adequately assured for most systems. The Human Factors evaluator must remember simply that he must adhere as closely to good experimental procedure as is possible under field conditions. This entails the establishment of a test plan which incorporates measuring performance under standard test conditions over enough test runs to be assured that an accurate estimate of the true performance of the operator under those conditions is obtained. Since the equipment evaluator must be concerned with the same problems the planning of the test runs to meet the needs of both the equipment and the operator evaluator will be possible for much of the testing.

Where control over test conditions is not possible in the field, the systematic observation and the measurement of operator performance will serve to identify those measurement points at which extreme variability in performance is occurring. These tasks may then be examined to determine whether (1) an obvious equipment design or procedural problem is present or (2) the measurement of performance at this point might be profitably undertaken in another test situation such as the simulator, test range or laboratory.

5.7.2 The Use of the Simulator for Test and Evaluation

In using the simulator as an evaluation tool the evaluator must be particularly aware of its limitations with respect to its representing the total system and its conditions of operation. On the one hand the simulator may not incorporate many of the tasks and conditions of the actual operating system. On the other, it allows for the repetition of standard test procedures and the introduction of variables and conditions into the test situation which might not be feasible or safe in the actual system. Therefore, the informed use of the dynamic simulator of the system for obtaining operator performance data during repeated runs of the mission, or segments of the mission, must be considered a valuable evaluation method.

Two test plans were used by the present authors in field tests to investigate the use of a systems trainer as an operator performance evaluation tool. The particular trainer used in this instance was the P-3C Weapon System Trainer (NAVTRADEVCEN Device 2F87). In the conduct of these tests two different lengths of test run were used. The first test employed short segments of the mission so that many repetitions of the run could be carried out during a test period. Objections to the short mission segment were voiced by the experienced operators used as subjects for the test. They expressed a lack of interest in repetition of the short mission segment and were not able to fully appreciate the evaluators need to obtain an adequate sample of performance scores. The second test of the simulator as an evaluation tool employed a much more inclusive set of conditions representing a major portion of an ASW mission. These test conditions were much more acceptable to the subject operators than was the short test.

The length of the test run and the number of variables and conditions to be included must be decided by the evaluator on the basis of his knowledge of good test conditions, his experience in dealing with the constraints of the field test situation including the limitations of the trainer, and his ability to interest and motivate the experienced operators performing the task. He should make it a point to enlist the operators' aid by informing them of the purpose of the tests and assuring them that they are not being evaluated as individuals - rather they are assisting in the evaluation of the system.

5.7.3 Test Ranges as Evaluation Tool

Certain tasks may be performed on test ranges under co[illegible] which can be repeated and performance data accumulated. In such cases the actual operating system may be used but only a certain segment of the mission be carried out. Data obtained under such test conditions can be collected under quite representative test conditions using measurement equipment which provides highly reliable data.

As an example of the use of ranges, the visual "mark-on-top" task of the P-3C pilot is a task in which accuracy of performance might well have been tested on an instrumented range as a single task with empirical performance data collected. The error distributions obtained under such conditions may be used to determine the contribution of measured pilot error to overall system error. The data obtained during repeated runs on the test range using representative operators can often give a more reliable estimate of operator performance than would be possible if the total mission test were used.

Test range data as well as the laboratory test data discussed below should play a large part in the evaluative decisions made during the development phase of the system. However, the field evaluator may find both the test range and the laboratory very useful. Usually he will have the advantage of greater knowledge of the tactical employment of the system as delivered and access to more representative operators with which to test the system. The data he collects will, therefore, reflect a more accurate picture of operational performance than will the earlier development data.

5.7.4 Laboratory Tests

As with the simulator and range tests, for some tasks it may be more appropriate to test the performance of the operator under quite closely controlled test conditions. The operator's ability to detect and respond to certain signals or inputs may be best tested in a controlled situation under the assumption that if he cannot perform satisfactorily under such conditions the probabilities are against his doing so in the complex operational system. The identification of such problem areas will come from operation of the actual system. In the collection of definitive data regarding them the laboratory test situation may often be appropriate.

5.8 WORKLOAD ANALYSIS

An important evaluative question with respect to man's performance in the system is whether he has the time necessary to complete the actions and decisions required of him throughout the course of the mission. What is his workload and is it excessive at any time?

During the development of the system the operator's workload will have been considered and assessed either formally and objectively or by subjective estimate. If formally assessed, the method used will most probably have employed the Operational Sequence Diagram or Mission Time Line to detail the tasks; have estimated the time required to execute each of the tasks; determined time available for completion of each of the several segments of the mission; and compared time required to time available as a quantitative estimate of workload. At the field evaluation level the evaluator must, through observation of operational performance, determine whether the tasks required during any mission segment are performed in their proper sequence within the time available. Falling behind in the execution of tasks or omitting tasks from the sequence indicate an excessive workload on the operator. When such conditions are observed one of the first questions the evaluator should consider is whether the experience of the operator being observed is equivalent to that of the operational system operator. It may be possible to conduct repeated runs of the particular mission segment in the simulator to determine the level of learning, learning curve and asymptotic performance of the operators. He can also gain an estimate of the effect of fatigue upon the performance and timely execution of the task.

5.9 EVALUATION OF THE DESIGN FOR OPERATOR FEEDBACK

For most control actions carried out during the operation of a system the results of the action are readily apparent to the operator. That is to say that there is immediate feedback or knowledge of results of his inputs to the system. However, the system may be so designed that, following an operator action, a lengthy sequence of system events unfolds before it is apparent to him whether or not the action was correct. His action may be one which commits the system to a given tactic and, if incorrect, will lead him to proceed on the assumption that the system is operating in a certain way when, in fact, it is not. By the time the result of the action error is apparent the mission may be irretrievably compromised.

The Human Factors evaluator should examine the Operational Sequence Diagram (or the Mission Time Line if the task analysis is in that form) and for each operator input to the system ask the following questions: (1) Does the operator receive feedback as to the accuracy of his actions? (2) If he receives feedback, does he receive it rapidly enough to correct an error before the mission is seriously compromised? If the operator receives no feedback, if it is delayed or if it is at all ambiguous a human factors design problem exists. Early examination of the task analysis to identify such problems may well save discovery of a design deficiency before costly performance measurement has been undertaken.

5.10 THE PERSISTENCE OF EARLIER PROCEDURES AND HABITS

If the evaluator observes the operator using procedures and methods carried over from an earlier system but not designed to be used witn the new he should inquire further into the matter. Almost every system is an evolutionary step from some predecessor. Personnel chosen to operate the new system during its evaluation will be those who are highly skilled with the ancestral system. If they persist in using habits and methods appropriate to the earlier system in performing with the new, a design problem may well be indicated. The evaluator should understand that the numan being has a proclivity for carrying old habits over into new situations. However, when old methods are persistently used in conflict with those called for by the design, the probability of a design fault should be investigated.

6.0 SUMMARY RECOMMENDATIONS

The recommendations in this report are directed at the performance evaluation of the human operator during field test and evaluation. The evaluation procedures dealt with here do not include the large amount of work carried out in evaluating the environmental variables and their effect upon the human operator.

The recommendations are based upon intensive study of a particular rather complex system being taken into the Navy inventory, the P-3C ASW system, and upon a less exhaustive study of the A-7A. They are based also upon first hand observations of the field evaluation process as it existed for these systems - again with emphasis upon the P-3C. The limitations upon the recommendations resulting from these conditions should be noted. At the same time, in arriving at these recommendations, the authors have applied the results of their experience in working with the design, development and evaluation of other military systems as well as upon the published findings and informal communications of others with experience with the problem.

6.1 MANAGEMENT AND PERSONNEL

The first and perhaps most important recommendation to be made would help greatly in alleviating the present constraints on effective human factors field evaluation. This is that serious consideration be given to the recommendations contained in Section 4.0 of this report. In that Section the need for assignment to the program office of personnel with authority and means for accomplishing human factors requirements is stated, recommendations for other personnel assignments during system development and test are made and training and cross training suggestions are outlined. The prime initial requirement for successful human factors field evaluation is the assignment at the program direction level of a qualified manager of human factors design and evaluation with authority and budget adequate to the task.

6.2 IDENTIFICATION OF MEASUREMENT POINTS

It is important that a detailed listing of the operator tasks and an Operational Sequence Diagram be generated and kept current throughout the development and test of the system. Both manual performance and tactical decision measurement points should be identified early in the development phase along with scales of measurement and criteria for acceptable performance. This information is then available in planning the field test and evaluation.

The evaluator will usually find it necessary to modify the task listing and OSD to reflect changes in the system just prior to and

during the test and evaluation process. It is recommended that he use the OSD to obtain, from either experienced operators of the system or Human Factors engineers familiar with the system, identifications of the points of human performance measurement, scales of measurement and criterion limits. This procedure is discussed in Section 5.4.7 of the report.

6.3 TESTING CONDITIONS

The Human Factors evaluator must be oriented toward obtaining reliable measures under test conditions which will allow him to forecast how the operator will perform with the complete system in the operational theater. The requirement for obtaining reliable measures and that of testing under the most representative conditions often may conflict. When test conditions are most like those of the operational theater the less likely it will be that conditions of test can be held constant from run to run and that reliable measures can be obtained.

In striking a balance between representative test conditions and reliability of data the field evaluator must strive to attain the most representative conditions possible while maintaining good test procedure. Since similarity to operational conditions is most nearly attained at the field test stage, test runs designed solely for control and measurement of human performance variables are necessary and must be scheduled. At the same time the use of the dynamic simulator, test ranges and laboratory-like settings should not be ruled out as a part of the field test procedure. The field evaluator who is familiar with the system and the field test constraints and who has available experienced operators to perform the tasks often can use these test conditions to advantage. This is particularly true when safety considerations or lack of control over test conditions limits the utility of the full system as a test situation. This subject is discussed in Section 5.7.

6.4 CRITERIA

The field evaluator's task is to obtain reliable performance data which he can compare to the criterion requirements for successful performance and reach an evaluative judgment as to the adequacy of design. Measurements must be taken which allow a direct comparison of performance with the level of performance required.

When specific criteria are not available the evaluator must obtain actual performance data over a series of runs upon which to base his predictions of whether the performance will significantly affect the successful performance of the mission by the total system. The sampling conditions under which he collects these predictive data are therefore critical to his predictions. The problem of performance criteria is discussed more fully in Section 5.6.3.

6.5 OPERATOR WORKLOAD

The evaluation of human operator performance may be simply stated as the assessment of the accuracy and timeliness with which he (1) processes data at the man-equipment interface and (2) makes the tactical decisions necessary for the system to accomplish its mission. Although assessments of workload will have been made earlier during the development of the system, the determination of whether the operator can successfully accomplish a sequence of tasks in the time available for accomplishing it within the actual mission must be determined under conditions as nearly representative of the operational mission as possible. Tendencies to get behind the task requirements or to substitute other procedures should be investigated for possible design inadequacies under the assumption that the stress of actual operational theater operations will tend to worsen operator performance.

6.6 PERSISTENCE OF FORMER PROCEDURES AND HABITS

It will be particularly important to the Human Factors evaluator to set up observational and/or report procedures for determining the degree to which deviations from design procedures occur through use of earlier learned habits. These pose a particular problem because of the possibility of the reversion to these procedures and habits during critical and stressful situations. This will be more likely to occur when the information displays are essentially the same as formerly used and the responses to the information (actions and procedures) have been changed.

7.0 BIBLIOGRAPHY AND SOURCE MATERIAL

SYSTEMS ANALYSIS AND INFORMATION REQUIREMENTS

Havron, M. D. & Jenkins, J. P. Information requirements methods as applied to development of displays and consoles. Engineering Psychology Branch, Psychological Sciences Division, Office of Naval Research, HSR-RR-61/4Sm, March 1961.

Hitchcock, L., Jr. The analysis of human performance within the operational flight environment. In Assessment of Skill and Performance in Flying, AGARD Conference Proceedings, No. 14, September 1966.

Kurke, M. I. Operational concept analysis and sources of field data. Human Factors, 1965, 7 (6).

McGrath, J. E. & Nordlie, P. G. Synthesis and comparison of system research methods. Human Sciences Research, Inc., Arlington, Va., Report Number 9, Contract Nonr 2425-(00), February 1960.

Meister, D. & Rabideau, G. F. Human Factors Evaluation in System Development. New York: Wiley, 1965.

Porter, E. H. A paradigm for system analysis of command and control functions. System Development Corporation, Santa Monica, Calif., October 1964.

Wilson, C. L. On-the-job and operational criteria. In R. Glaser (Ed.), Training Research and Education. Pittsburg: U. of Pittsburg Press, 1962. Pp. 410-412.

Wilson, D. A. A procedural guide to operational sequence analysis in personnel research. SRM 66-19, U. S. Naval Personnel Research Activity, San Diego, California, January 1966 (AD 477 767L).

Wilson, D. A. Application of automatic data processing techniques to task analysis diagramming. SRM 68-8, U. S. Naval Personnel Research Activity, San Diego, California, October 1967, (AD 660 002).

Wilson, D. A. A procedural guide to an automatic data processing method of task analysis diagramming. SRM 69-3, U. S. Naval Personnel Research Activity, San Diego, California, August 1968, (AD 677 794).

HUMAN ENGINEERING CHECKLISTS

Cornell Aeronautical Laboratory, Inc. Pocket data for human factors engineering. June 1959.

Fitzpatrick, R. A checklist for human factors in operational suitability tests. American Institute for Research, Contract AF 24(601)-124, Project No. 7800, Research Directorate, Air Force Special Weapons Center, Air Research and Development Command, Kirtland AFB, New Mexico, August 1955.

Greek, D. C. Checklist of human engineering design principles. MD58-334, North American Aviation, Inc., Missile Division, January 1959.

Meister, D. & Rabideau, G. F. Human factors evaluation in system development. New York: Wiley, 1965.

Philco-Ford Corporation, Human engineering design checklist. WDL-TRI-11968A, Philco-Ford Corporation, WDL Division, Palo Alto, California, 8 May 1964.

Plath, D. W. Human engineering checklist. Autonetics, A Division of North American Aviation, Inc., January 1966.

PERFORMANCE EVALUATIONS AND SYSTEM EFFECTIVENESS

Adams, J. A. & McAbee, W. H. A program for evaluation of human factors in category II testing of air weapons control system 412L (phase II configuration). Air Force Systems Command, Eglin Air Force Base, Florida, May 1962.

Air Force Ballistic Missile Division (ARDC), Personnel subsystem testing for ballistic missile and space systems. AFBMD Exhibit 60-1, April 1960.

Barfoot, C. B. A preliminary cost-effectiveness handbook. Headquarters, United States Army Combat Developments Command, Ft. Belvoir, Va., 15 December 1963.

Barrett, G. V., Thornton, C. L. & Cabe, P. A. Human factors evaluation of a computer based information storage and retrieval system. Human Factors, 1968, 10 (4).

Borchers, H. Are ballistic missile test programs structured to support adequate evaluation of human performance? Human Factors, 1964, 6 (6).

Borden, J. & McGrath, J. Geographic orientation in aircraft pilots: Field validation of a post-flight method of reporting navigation performance. JANAIR Report 680714, Human Factors Research, Inc., Santa Barbara Research Park, Goleta, California, July 1968.

Bowen, H. M, Bishop, E. W. & Promisel, D. Study, assessment of pilot proficiency. U. S. Naval Training Device Center, Port Washington, N. Y., Technical Report: NAVTRADEVCEN 1614-1, August 1966.

Brictson, C. A. Measures of pilot performance: Comparative analysis of day and night carrier recoveries. Dunlap and Associates, Inc., Santa Monica, California, June 1966.

Chapin, P. W., Detrick, R. E. & Ilg, R. P. Flying qualities and performance evaluation of the A-7A airplane. Report No. NATC-FT-35R-68, May 1968, Naval Air Test Center, Patuxent River, Maryland.

Cooper, L. J. & Manion, R. C. Testing the OPTEVFOR Guidebook for the collection of human factors data, Report I: Direct Observation. Report 4040F-1, State College, Penn., HRB-Singer, Inc., Nov. 15, 1967, (Preliminary Report).

Curtin, J. G., Dougherty, D. J. & Emery, J. H. Evaluation of the RH-2 cockpit mock-up. Bell Helicopter Company, Fort Worth, Texas, August 1965.

Data Design Laboratories, Human factors prediction evaluation. Data-Design Laboratories for the Bureau of Naval Personnel, Contract Nonr 4739(00), August 1967.

Glaser, R. Evaluation of personnel proficiency. In J. D. Folley, Jr., (Ed.), Human Factors Methods for Systems Design. AIR-290-60-FR-225, American Institute for Research, 1960. AD 232 646.

Gross, R. L., Lindquist, O. H. Peterson, J. R. & Blanchard, J. W. The application, validation and automation of a method for delineating and quantifying aerospace flight crew performance. In Symposium and Workshop in the Quantification of Human Performance. Albuquerque, U. of New Mexico, 1964.

Hagen, W. C. Techniques for the allocation and evaluation of human resources. Martin Marietta Corporation, Orlando, Florida, March 1967, OR 8735.

Houston, R. C., Smith, J. F. & Flexman, R. E. Performance of student pilots flying the T-6 aircraft in primary pilot training. 1954, AFPTRC-TR-54-109.

Jeantheau, G. G. & Anderson, B. G. Training system use and effectiveness evaluation. U. S. Naval Training Device Center, Port Washington, N. Y., Technical Report: NAVTRADEVCEN 1743-1, June 1966.

Johnson, G. W. Navy laboratory role in human performance effectiveness. In Proceedings, Symposium on Human Performance Quantifications in Systems Effectiveness. Naval Material Command and the National Academy of Engineering, Washington, D. C., January 1967, AD 650-791.

Keenan, J. J., Parker, T. C. & Lenzychi, H. P. Concepts and practices in the assessment of human performance in Air Force systems. AMRL-TR-65-168, Aerospace Medical Research Laboratories, AFSC, Wright-Patterson AFB, Ohio, September 1965.

Kryter, K. D. Proposed methods for the calculation of the articulation index. ESD TDR-62-5, Operational Applications Office, Electronics Systems Division, Bedford, Mass., October 1961.

Kryter, K. D., Flanagan, G. & Williams, C. A test of the 20-band and octave-band methods of computing the articulation index. ESD-TRD-62-4, Operational Applications Laboratory, Electronic Systems Division, Bedford, Massachusetts, October 1961.

Kryter, K. D. & Ball, J. H. SCIM -- A meter for measuring the performance of speech communications systems. TDR Report No. ESD-TDR-64-674, Electronic Systems Division, Air Force Systems Command, October 1964.

Landis, D., Slivka, R. M., Jones, J. M. & Silver, C. A. Experiments in display evaluation, Phase I report. Technical Report I-194, The Franklin Institute Research Laboratories, Contract Nonr-4832, July 1967.

Leuba, H. R. & Boteilho, R. J. Evaluation of system design adequacy. ARINC Research Corporation, Washington, D. C., 1963.

Luehrs, R. E., Capt. Human error research and analysis program (HERAP). In Assessment of Skill and Performance in Flying, AGARD Conference Proceedings No. 14, September 1966.

Lytton, L. E. Evaluation of a vertical-scale, fixed-index instrument display panel for the X-15 airplane. NASA TN D-3967, Washington, D. C., May 1967.

McKendry, J. M. & Harrison, P. W. Assessing human factors requirements in the test and evaluation stage of systems development, Vol. I. State College, Pa., HRB-Singer, Inc., June 1964.

McKendry, J. M. & Harrison, P. C. Assessing human factors requirements in the test and evaluation stage of systems development, Vol. II. State College, Pa., HRB-Singer, Inc., June 1964.

Mahler, W. R. & Bennett, G. K. Psychological studies of advanced Naval air training: Evaluation of operational flight trainers. New York: Psychological Corp., Sept. 1950.

Manion, R. C., Cooper, L. J. & Thomas, D. B. Testing OPTEVFOR guidebook for the collection of human factors data, Report 2: The questionnaire. State College, Pa., HRB-Singer, Inc., November 1967 (Report 4040F-3).

Manion, R. C. & Cooper, L. J. Testing the OPTEVFOR guidebook for the collection of human factors data, Report 3: The checklist. State College, Pa., HRB-Singer, Inc., November 1967, (Report 4040-F-2).

Manion, R. C. & Cooper, L. J. Testing the OPTEVFOR guidebook for the collection of human factors data, Report 4: The interview. State College, Pa., HRB-Singer, Inc., November 1967, (Report 4040-F-4).

Markel, G. A. Toward a general methodology for systems evaluation. State College, Pa., HRB-Singer, Inc., July 1965.

Meister, D. The measurement of man-machine systems under field operational conditions. Paper presented at the ONR Ninth Annual Human Engineering Conference, St. Louis, Missouri, June 1-2, 1961.

Meister, D. The measurement of man-machine systems under field operational conditions. Human Factors, 1964, 3 (4).

Meister, D. Methods of predicting human reliability in man-machine systems. Human Factors, 1964, 6 (6).

Meister, D. & Rabideau, G. F. Human factors evaluation in system development. New York: Wiley, 1965.

Newell, F. D. & Pietrzak, P. E. In-flight measurement of human response characteristics. AIAA Flight Test, Simulation and Support Conference, Cocoa Beach, Florida, February 6-8, 1967.

Obermayer, R. W. Simulation, models and games: Sources of measurement. Human Factors, 1964, 6 (6).

Pickrel, E. W. & McDonald, T. A. Quantification of human performance in large, complex systems. Human Factors, 1964, 6 (6).

Rabideau, G. F. Field measurement of human performance in man-machine systems. Human Factors, 1964, 6 (6)

Rappaport, M. Further considerations about the what and how of field testing. Unpublished report, 1961.

Rook, L. W., Jr. A method for evaluating the human error contribution to system degradation. Human Error Quantification, A Symposium, A. D. Swain, Chairman, SCR-610, Sandia Corporation, Albuquerque, New Mexico, April 1963.

Rook, L. W., Jr. Evaluation of system performance from rank-order data. Human Factors, 1964, 6 (5).

Schum, D. A., Elam, C. B. & Matheny, W. G. The static method of display design evaluation: With special reference to the altimeter. Bell Helicopter Company Report, Contract AF33(616)-8236, Aeronautical Systems Command, Wright-Patterson AFB, Ohio, January, 1962.

Shearer, J. W., Peterson, D. A. & Slebodnick, E. B. Techniques for human factors evaluation of prototype special weapons and associated equipment. American Institute for Research, Contract AF 29(601)-513, Air Research and Development Command, Kirtland AFB, New Mexico, 1959.

Sheldon, M. S. & Zagorski, H. J. Project NORM field evaluation test report. System Development Corporation, Santa Monica, California, July 1966.

Siegel, A. I. & Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: A description of a model and the results of its first application. Applied Psychological Services (ONR report) Feb. 1959.

Siegel, A. I. & Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: Application of a previously derived model to the launching of an air-to-air missile. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1959.

Siegel, A. I., Wolf, J. J. & Crain, K. Techniques for evaluating operator loading in man-machine systems: A "model" for digital simulation of one and two-operator man-machine systems. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, March 1961.

Siegel, A. R. & Wolf, J. J. A technique for evaluating man-machine system designs. Human Factors, 1961, 3 (1).

Siegel, A. I. & Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: A further application of a "model" for digital simulation of one or two-operator man-machine systems. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1961.

Siegel, A. I., Wolf, J. J. & Sorenson, R. T. Techniques for evaluating operator loading in man-machine systems: Evaluation of a one or a two-operator system evaluative model through a controlled laboratory test. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, July 1962.

Swain, A. D. Some problems in the measurement of human performance in man-machine systems. Human Factors, 1964, 6 (6).

Smode, A. F., Gruber, A., & Ely, J. H. The measurement of advanced flight vehicle crew proficiency in synthetic ground environments. Report MRL-TDR-62-2, Aerospace Medical Division, Wright-Patterson AFB, February 1962.

Ware, C. T., Jr. Individual and situational variables affecting human performance. Human Factors, 1964, 6 (6).

Widrewitz, J. Concepts relative to system effectiveness. Rome Air Development Center, Air Force Systems Command, Griffiss Air Force Base, N. Y., June 1964.

William, C. E., Hecker, Michael, H. L. & Kryter, K. D. Methods for psychoacoustic evaluation of speech communication systems. Technical Documentary Report No. ESD-TDR-65-153, Electronic Systems Division, Air Force Systems Command, December 1964.

Wilson, C. L. On-the-job and operational criteria. In R. Glaser (Ed.), Training Research and Education. Pittsburgh: U. of Pittsburg Press, 1962. Pp. 410-412.

MODELS AND METHODS

Cagley, E. M. A general network simulation model. NREC Tech. Report 31, U. S. Engineer Mathematical Computational Agency, April 1964.

Cochran, W. G. Sampling techniques, (2nd ed.) New York: Wiley, 1964.

Guilford, J. P. Psychometric methods, (2nd ed.) New York: McGraw-Hill, 1954.

Hendricks, J. W. A system for development of mathematical models. U. S. Naval Radiological Defense Laboratory, San Francisco, California, USNRDL-TR-749, April 1964.

Johnson, P. O. Statistical methods in research. New York: Prentice-Hall, 1949.

Leuba, H. R. Quantification in man-machine systems. Human Factors, 1964, 6 (6).

Myers, L. B., Carter, R. G. & Hostetter, R. S. Guidebook for the collection of human factors data (Second Printing), State College, Pa., HRB-Singer, Inc. Contract Nonr 4751 (00), June 1966.

Rook, L. W., Jr. Evaluation of system performance from rank-order data. Human Factors, 1964, 6 (5).

Ryan, P. W. Cockpit geometry evaluation, Vol. 1-4. Boeing Co. (Prepared for JANAIR Program) February 1970.

Siegel, A. I. & Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: A description of a model and the results of its first application. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, February 1959.

Siegel, A. I. & Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: Application of a previously derived model to the launching of an air-to-air missile. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1959.

Siegel, A. I., Wolf, J. J. & Crain, Jr. Techniques for evaluating operator loading in man-machine systems: A "model" for digital simulation of one and two-operator man-machine systems. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, March 1961.

Siegel, A. I. & Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: A further application of a "model" for digital simulation of one or two-operator man-machine systems. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1961.

Siegel, A. I., Wolf, J. J. & Sorenson, R. T. Techniques for evaluating operator loading in man-machine systems: Evaluation of a one or a two-operator system evaluative model through a controlled laboratory test. Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, July 1962.

Torgerson, W. S. Theory and methods of scaling, (Second Printing) New York: Wiley, 1960.

DATA STORES (BANK)

Barrett, G. V., Thornton, C. L. & Cabe, P. A. Human factors evaluation of a computer based information storage and retrieval system. Human Factors, 1968, 10 (4).

RELEVANT SPECIFICATIONS, STANDARDS AND INSTRUCTIONS

Department of Defense Directive Number 5000.1, July 13, 1971.
Subject: Acquisition of Major Defense Systems.

MIL-H-46855	- Human Engineering Requirements for Military Systems, Equipment and Facilities
MIL-L-5667	- Lighting Equipment, Aircraft Instrument Panel, General Specification for Installation of
MIL-M-18012	- Markings for Aircrew Station Displays Design and Configuration of
MIL-STD-203	- Aircrew Station Controls and Displays for Fixed Wing Aircraft
MIL-STD-411	- Aircrew Station Signals
MIL-STD-783	- Nomenclature and Abbreviations in Aircrew Stations
MIL-STD-850	- Aircrew Station Vision Requirements for Military Aircraft
MIL-STD-1472	- Human Engineering Design Criteria for Military Systems, Equipment and Facilities

APPENDIX A

SUGGESTED OUTLINE FOR COURSE IN

HUMAN FACTORS TEST AND EVALUATION

Note: Material suggested in this outline comes in part from Smode, et al., 1962. The use of that report is suggested as a suitable part of the text material for the course outlined in this Appendix.

APPENDIX A

This appendix outlines a course of study designed for field test personnel whose particular interest and concern is human factors testing. The course is intended to cover items essential to an overall orientation in human factors test and evaluation and to provide methods and techniques necessary to effective evaluation.

In the outline a differentiation is made between field research and field evaluation with similarities and differences drawn between them. While the primary aim of the course is to provide guidance for field evaluation, points relevant to field research are given in order that the human factors evaluator may have some guidance in field research techniques should the opportunity arise to apply them during a field evaluation.

COURSE OUTLINE

1.0 Distinction between field research and field evaluation.

1.1 Field Research

1.1.1 More opportunity to identify and control the parameters and variables of interest.

1.1.2 Usually collecting data to establish norms or make comparisons.

1.1.3 More opportunity to introduce special instrumentation to obtain performance measures.

1.1.4 More flexibility in changing procedures and equipment as testing progresses to achieve the desired goals of testing.

1.1.5 May be testing to determine how well system, subsystem or component meets some set or required level of performance.

1.2 Field Evaluation

1.2.1 Nearly always testing to determine whether system performance meets some specified level of design performance.

1.2.2 Nearly always there is a limitation on time within which the evaluation is to be performed.

1.2.3 Often must fit measures of man's performance into the tests of equipment.

1.2.4 Little or no opportunity to vary independent variables systematically.

2.0 General setting within which field research and evaluation are conducted.

2.1 Some fixed time span.

2.2 Testing both equipment and men.

2.2.1 Some tests peculiar to equipment alone, e.g., how it functions under the field conditions.

2.2.2 Some tests peculiar to man alone, e.g., his physiological state under the field conditions.

2.2.3 Some tests peculiar to the interaction between men and equipment, e.g., how well man can operate or maintain equipment under the field conditions.

2.3 Human factors testing often goes "piggy back" on equipment testing, i.e., must be fitted in and around equipment testing. This is a fact of human factors testing which must be recognized in setting up human factors research and evaluations in the field. A great deal of human factors data collecting can be carried out in conjunction and simultaneous with equipment testing. However, it will be necessary to program specific blocks of time for collecting human factors data independent of equipment tests.

3.0 Requirement for thorough knowledge of the system under test, its operation and its operating environment.

3.1 To isolate and define the important parameters which may influence performance.

3.2 To set up methods of either controlling or systematically varying these parameters.

3.3 To determine what measures are appropriate and at what test points they will be taken.

4.0 Methods for determining necessary system details.

4.1 Data flow analysis.

4.1.1 What inputs are (or must be) received by each component or subsystem and what outputs are (or must be) made to the next component(s) or subsystem(s).

4.1.2 For man as an information processor we are interested in what information he must receive or is the system designed for him to receive, what transformations of the information he must make, and what "informational" outputs he must make to other components of the system.

4.1.3 A number of formats with their own nomenclature or symbology have been developed. Each has its own particular assets.

4.2 Time Line Analysis - serves both as a means of learning the system and evaluating workload.

4.3 Operational Sequence Diagram - serves as means of learning the system and identification of manual performance and tactical decision measurement points.

5.0 Some general principles of sound research and evaluation procedure for which to aim.

5.1 Standardization.

5.1.1 Test conditions.

5.1.2 Performance measures.

5.1.3 Observers.

5.1.4 Environmental effects.

5.2 Control or assessment of relevant variables.

5.2.1 If variable cannot be controlled it should be measured at the time dependent variables are measured.

5.3 Explicit statement of independent and dependent variables with no variation in their meaning or method of measurement.

5.4 Dependent variables.

5.4.1 These are, in effect, precise statements of what we want to know about how the system or the man within the system performs.

5.4.2 Must order dependent variables in order of importance since resources and/or time are usually limited.

5.4.3 Must decide upon points in system at which measurements will be taken.

5.4.4 Must determine the method of measurement, i.e., direct recording, direct observation, rating scales, questionnaire, etc.

5.4.5 Always detail in advance the method of data reduction, analysis and presentation. Shotgun approach to data collection is not feasible in field situations.

5.5 General classes of dependent variables.

5.5.1 Man's outputs.

5.5.1.1 Time to perform.

5.5.1.2 Accuracy of performance.

5.5.2 Man's inputs.

5.5.2.1 Control displacements or forces.

5.5.3 Man's physiological state.

5.6 Independent Variables.

5.6.1 In field research may have the opportunity to assess, control or vary systematically the independent variables. In field evaluation usually have opportunity only to assess these variables.

5.6.2 From system and task analysis identify and describe both system and environmental parameters likely to affect performance.

5.6.3 Define and specify how independent variable is measured and over what range it will be varied.

6.0 Criteria

6.1 Implies some value judgment as to the "goodness" of the performance.

6.2 Measurement per se does not provide value judgments.

6.3 These value judgments must be expressed in terms of the defined purpose or mission of the system.

6.4 Ultimate vs. actual criteria.

6.4.1 Seldom possible to obtain direct measures of the ultimate criteria.

6.4.2 Usually necessary to select some actual (intermediate) criteria.

6.4.3 Must then use these actual (intermediate) criteria in evaluating performance.

6.4.4 There is no certain method for specifying the actual criteria.

6.4.5 Sources of error in selecting actual criteria.

6.4.5.1 Unreliability.

6.4.5.2 Irrelevancy - the lack of relation to ultimate criterion.

6.4.5.3 Contamination - ingredients in the actual criteria which do not, in fact, exist in the ultimate criterion.

6.4.5.4 Distortion - errors arising from assigning incorrect weights to the separate factors that comprise the actual criteria.

6.5 Establishing valid criteria.

6.5.1 No established procedures.

6.5.2 Recognizing importance of criteria selection and types of errors which might be present are good starting points.

6.5.3 Steps which should lead to more useful and relevant criteria.

6.5.3.1 Define the activity - specify to extent possible the activity desired for successful and proficient performance.

6.5.3.2 Analyze the activity - consider the activity in terms of purposes or goals, behavior and skills involved, their relative importance and standards of performance expected.

6.5.3.3 Define successful performance.

6.5.3.4 Develop sub-criteria to measure each element of success.

6.5.3.5 As appropriate develop a combined measure of successful performance.

6.6 Combining criteria.

6.6.1 Often necessary that several criteria, all of which are relevant for a particular activity be used. In such case it may be desirable to combine them into a single comprehensive one.

6.6.2 Combining will usually involve assigning relative weights to the individual criteria.

6.6.3 Rules for combining criteria.

6.6.3.1 Weight in accordance with their relevance to the ultimate criterion.

6.6.3.2 Criteria which repeat or overlap factors in other criteria should receive low weight.

6.6.3.3 Other things being equal the more reliable criteria should receive more weight.

6.6.4 Caution must be exercised in applying weights to raw score values - use standard scores.

7.0 Measurement of performance.

7.1 What to measure.

7.1.1 Should be preceded by an explicit statement of the research questions being asked.

7.1.2 Need analysis of the system which allows identification of the points within the system at which performance can be measured and recorded.

7.1.3 Identification of critical tasks.

7.1.3.1 Usually not feasible to measure all points identified so must select for measurement those tasks on which good performance leads to mission success and poor performance leads to mission failure.

7.1.3.2 In identifying critical tasks asking the following questions with respect to the tasks is helpful.

Would below-minimum performance;

- lead to an accident?
- result directly in mission failure?
- be impossible to remedy within the time constraints or not at all?
- be difficult to detect because of inadequate information feedback?
- recur over time in such a way as to produce a cumulative effect?
- contribute a large proportion of time to the total time required for some larger and critical function?

7.2 Levels of measurement

7.2.1 Nominal scale.

7.2.2 Ordinal scale.

7.2.3 Interval scale.

7.2.4 Ratio scale.

7.3 Specificity of measures.

7.3.1 Over-all measures.

7.3.1.1 Global indices of sub-system or system performance associated with mission segments or complete mission.

7.3.1.2 Useful in assessment since it is descriptive of some end result which can be compared with the standard.

7.3.1.3 Weak in analytic sense since they provide no detailed information on performance beyond the outputs sampled.

7.3.2 Diagnostic measures.

7.3.2.1 Quite specific, identifying elements of job performance in specific skill areas.

7.3.2.2 Since they are concerned with smaller more precisely defined units of behavior they lend themselves more readily to objective measurement.

7.4 Accuracy of measurement.

7.4.1 Refers to how close the obtained value or measure is to the true value.

7.4.2 There is no single way to assure measurement accuracy. Accuracy may be improved by the following means.

7.4.2.1 Increase scope of measurement to be taken - include additional aspects of relevant behavior.

7.4.2.2 Increase the number of observations on which summary statistics, e.g., are based.

7.4.2.3 Control the conditions under which measurements are taken.

7.5 Reliability of measurement.

7.5.1 Definition - agreement or consistency of measures from repeated observations.

7.5.2 Relation between reliability and validity.

7.5.3 Absolute expression of reliability - standard error of measurement.

7.5.4 Relative measures of reliability - expressed in terms of correlation.

7.6 Validity of measurement.

7.6.1 Definition - degree to which measuring instruments measure what they are intended to measure.

7.6.2 Four types of validity.

7.6.2.1 Content validity - logical validity based on expert opinion or other logical considerations.

7.6.2.2 Concurrent validity - statistical validity - correlation with other task or dimension external to the measurement.

7.6.2.3 Predictive validity-statistical correlation between obtained measures and future states on some task or dimension external to the measurement.

7.6.2.4 Construct validity - logical validity - where the emphasis is on the trait, quality or ability presumed to underlie the measures being taken.

7.7 Objective (quantitative) vs. subjective (qualitative) measures.

7.7.1 Objective measures.

7.7.1.1 Generally permit measurement relatively independent of the observer.

7.7.1.2 Generally of higher reliability than subjective.

7.7.1.3 Greatest objectivity obtained by means of recording instruments where a permanent record of behavior is obtained at the time of occurrence.

7.7.1.4 Insistence upon complete objectivity tends to result in omission of a variety of critical job components because of inability to measure them objectively.

7.7.1.5 Can result in impractical gadgetry and procedures.

7.7.1.6 Relatively free from observer bias.

7.7.2 Subjective measures.

7.7.2.1 Generally dependent upon the characteristics of the observers - may introduce bias.

7.7.2.2 Inter-observer reliability not always high.

7.7.2.3 More flexibility in administration.

7.7.3 Ratings (a form of subjective measurement).

7.7.3.1 Rating procedures.

- Rating scales - rater makes judgment on scale of defined categories.
- Comparative systems - pair people or units with respect to each other.
- Check lists - judgments by raters as to which of a series of descriptive terms either are or are not applicable to the units being evaluated.
- Critical incidents - recording actual incidents as behaviors which are especially effective or ineffective in the accomplishment of the mission - Standing of a unit is indicated by frequency of occurrence of reported incidents.

7.7.3.2 The sources of bias in ratings.

- Halo effect.
- Leniency error.
- Error of central tendency.
- Contrast error - tendency to rate in opposite direction on a dimension from how the raters see themselves.
- Proximity error - tendency for ratings to be more related when made close to each other in time.

7.8 Individual vs. crew performance.

7.8.1 Crew performance must be regarded as more than the sum of the individual performances.

7.8.2 Measures of crew performance.

7.8.2.1 Synchronization of action.

7.8.2.2 Response improvisation.

7.8.2.3 Amount of time spent interacting - good crew should reduce individual interaction to a minimum so that more effort is devoted to the job and less to coordinating.

7.8.2.4 Amount of communication - the less the communication the higher the degree of coordination.

7.8.2.5 Freedom for interpersonal communications.

7.8.2.6 Monitoring and/or making some responses for another crew member.

7.8.2.7 Aiding in the detection of out-of-tolerance conditions.

7.8.2.8 Sharing of risk activities among crew members.

8.0 Procedural steps in assessment of performance.

8.1 Conduct thorough analysis of the tasks and generate Operational Sequence Diagram or Mission Time Line.

8.2 Select test points and measures appropriate to the behavior to be evaluated.

8.3 Define performance requirements of the task as appropriate.

8.4 Identify important and critical aspects of the task and the environment.

8.5 Determine conditions under which measures will be taken.

8.6 Determine techniques for obtaining measurement data and for combining measures as appropriate.

8.7 Specify methods of data analysis.

9.0 Subjects or operators upon which data is collected (subject sampling)

9.1 Sample size.

9.2 Sample composition with respect to experience and other factors and its relation to purposes of measurement and usefulness of the data.

10.0 Data collection and treatment.

10.1 Experimental methods.

10.1.1 Single variable design.

10.1.2 Multi-variate design.

10.1.2.1 Each subject his own control.

10.1.2.2 Independent groups.

10.2 Specific measures of performance.

10.2.1 Procedural tasks.

10.2.1.1 Time.

10.2.1.2 Accuracy.

10.2.2 Closed-loop tracking accuracy (compensatory and pursuit)

10.2.2.1 Integration of error

10.2.2.2 Number of crossings.

10.2.2.3 Time on target.

10.2.2.4 Frequency of catastrophic errors.

10.2.3 Operator output.

10.2.3.1 Power density spectrum.

10.2.3.2 Standard deviation about mean output.

10.2.4 Crew coordination.

10.2.4.1 Overall gross measure of crew output in terms of time to accomplish task and accuracy in accomplishing.

10.2.4.2 Number of communications between members.

10.2.5 Decision making tasks.

10.2.5.1 Time.

10.2.5.2 Accuracy.

10.2.6 Perceptual and motor skills.

10.2.6.1 Psychopnysical measures.

10.3 Parametric statistics.

10.3.1 Measures of central tendency.

10.3.2 Measures of deviation from standard (CE).

10.3.3 Measures of variability.

10.3.4 Measures of correlation.

10.3.5 Tests of reliability of differences.

10.4 Non-parametric statistics.

10.4.1 Difference from parametric statistics.

10.4.2 Tests of reliability of differences.

10.4.3 Test of correlation.

10.5 Presentation of results.

10.5.1 Pictorial graphs or charts.

10.5.2 Significance tables - differences in terms of likelihood function.

APPENDIX B

SAMPLE OPERATIONAL SEQUENCE DIAGRAM

CONSTRUCTION OF BEHAVIORAL CODE WORDS

FIRST LETTER		
HUMAN OR MACHINE	H	(Human)
	M	(Machine)
SECOND LETTER		
BEHAVIOR	A	(Act)
	D	(Decide)
	T	(Transmit)
	R	(Receive)
	S	(Store)
	P	(Use Previously Stored Information)
	M	(Monitor)
THIRD LETTER		
MEANS OF PERFORMANCE	S	(Speech)
	P	(Phone, Sound-power)
	I	(Intercom)
	E	(Electrical or Electronic)
	T	(Touch)
	V	(Visual)
	F	(Filed)
FOURTH LETTER		
DISPLAYED OR NOT	D	(Displayed)
	Blank	(Not Displayed)
FIFTH LETTER		
INVERSE MEANING *	G	(Go, yes, normal, etc.)
	N	(No-Go, no, abnormal, etc.)

* Although this symbol is not used in the illustrations to follow it may be used where necessary or for convenience to indicate the opposite or lack of an action, i.e., no-go. (See Wilson, D. A., 1966).

OPERATIONAL SEQUENCE DIAGRAM

	TACCO		COMPUTER
		(MRE)	
HRI	1. Receive Initial Contact Report (M-14)		
HRI	2. Acknowledge Contact Report		
		(MTE)	
HRT	3. Request Classification and Bearing Line		
		(MTE)	
		(MRE)	
HRT	4. Receive Acknowledgement of Request		
HA	5. Rescale MDD as Required		
		(MTE)	(MAE) Display Desired Scale
		(MRE)	
HRV	6. Observe MDD Rescaled		(MRF)
		(MRE)	(MPED) Assimilate NAV and Buoy RF Data and Display Location of M-14
HRV	7. Observe Location of M-14		
		(MRE)	
HRV	8. Observe Navigation Data		(MRF)
		(MRE)	(MPED) Assimilate NAV Data and Display Position of Aircraft Symbol
HRV	9. Observe Position of Aircraft Symbol		(MRF)
		(MRE)	(MPED) Assimilate NAV and Time Data and Display Change of Aircraft Symbol Position
HRV	10. Observe Direction that Aircraft Symbol is Moving		
HTT	11. Request Pilot Turn Toward M-14		
		(MTE)	
		(MRE)	
HRI	12. Receive Acknowledgement of Request		(MRF)
		(MRE)	(MPED) Assimilate NAV and Time Data and Display Aircraft Symbol Turning
HRV	13. Observe Aircraft Symbol Turning		
		(MRE)	
HRI	14. Receive Classification Report		

SYSTEM P-3C ASW AIRCRAFT
OPERATOR - TACCO
PHASE - SEARCH
EVENT - INITIAL CONTACT

PAGE 1 OF 10
REVISION NO 0
DATE 14 JUN 1971

TASK SEQUENCE		RELATED COMPUTER OPERATION	
	BEHAVIOR CODES	BEHAVIOR CODES	
		MRE MAED	DISPLAY CONTACT SYMBOL
	MRE	MTE	
OBSERVE CONTACT SYMBOL	HRV		
PRESS ACK CONTACT SWITCH	HAT MTE		
		MRE MAE	REMOVE CONTACT SYMBOL
	MRE	MTE	
OBSERVE CONTACT SYMBOL REMOVED	HRV		
		MRF MPED MTE	DISPLAY AIRCRAFT AND CONTACT BUOY SYMBOLS
	MRED		
OBSERVE POSITION OF AIRCRAFT RELATIVE TO CONTACT BUOY	HRV		
		MRF MPED	DISPLAY AIRCRAFT SYMBOL MOVEMENT
	MRED	MTE	
OBSERVE DIRECTION OF AIRCRAFT MOVEMENT	HRV		
REQUEST PILOT TO BEGIN TURN TO STARBOARD	HTI MTE		
MOVE HOOK SYMBOL TO FTP EI	HAT MTE	MRE	
		MAED	DISPLAY HOOK SYMBOL AT FTP EI
	MRE	MTE	
OBSERVE HOOK SYMBOL AT FTP EI	HRV		
PRESS HOOK VERIFY SWITCH	HAT MTE		
		MRE MAED	DISPLAY FLASHING SYMBOL
	MRE	MTE	
OBSERVE FLASHING SYMBOL	HRV		
PRESS DESTROY PT DATA	HAT		

APPENDIX C

ILLUSTRATION OF OPERATIONAL SEQUENCE DIAGRAM AND INSTRUCTIONS FOR IDENTIFYING MEASUREMENT POINTS, TYPE OF DATA WHICH MAY BE COLLECTED AND CRITERION LEVEL OF ACCEPTANCE PERFORMANCE

ILLUSTRATIVE QUESTIONNAIRE USED FOR DATA COLLECTION IN CONNECTION WITH ASW TACTICAL COORDINATOR (TACCO)

Name: ____________________

Rank: ____________________

TACCO Experience: P-3C (Place a check in the appropriate space)

[ ] Recently Qualified: [ ] 1 - 3 Mo.; [ ] 3 - 6 Mo.; [ ] 6 - 9 Mo.:

[ ] 9 - 12 Mo.; [ ] More than 12 Mo.

TACCO Experience: Other Aircraft

[ ] None; [ ] less than 1 Yr.; [ ] 1 Yr.; [ ] 2 Yr.: [ ] 3 Yr.;

[ ] 4 Yr.; [ ] 5 Yr. or More

GENERAL INFORMATION

This questionnaire was prepared to gain information about the role of the TACCO in the P-3C ASW weapon system. The information will be used as part of a program for developing evaluation techniques to be used in human factors field evaluation. Your careful consideration in answering the questions below will be valuable in the proper development of these evaluation techniques.

The TACCO's crew station was selected because the system has been changed somewhat from previous ASW systems due to the use of onboard computers. Mission success depends on system design and how well the equipment is designed for the TACCO's use. Your help will be significant because the information that you give will point to important factors which will aid in system design evaluation.

In order to establish a baseline of data for reference it is necessary to consider that situation in which all the components in the system are functioning correctly. Listed below is a sequence of tasks that are required between the time that contact is made on one buoy in a barrier pattern and the time that the bias velocity is accepted, rejected or modified. It is recognized that each TACCO would probably have a task sequence that would vary somewhat from the sequence presented here, so just consider that this is only one of many possible solutions.

The conditions at the time that this task sequence begins are as follows:

(a) You have a three row barrier pattern.
(b) Contact has been made on one buoy.

INTRODUCTION

There are a number of times on an ASW mission during which you (1) receive information from the system, (2) interpret the information, and (3) put information into the system. These points are potential error producers. While it is understood that each task is important, there are some points in the sequence of tasks at which measurements could be taken. We cannot directly record the information you receive or how you interpret the information. That is, we cannot directly record what you read a pointer position on a dial to be. We would have to interrupt your work and ask you. Nor can we directly record how you interpreted the information you received. However, we can measure your outputs into the system in the form of the actions you make and the things you report.

INSTRUCTIONS

Using the task and TACCO scope illustrations, select the points at which you input information to the system and at which you feel measurements of performance could be obtained. After you have selected a task or task sequence as a measurement point, indicate (1) what type of data you feel can be taken at that point, and (2) whether there is an allowable performance envelope around the measurement point.

Please mark with an asterisk (*) by each task or task sequence that is selected as a measurement point. Use the space to the right to identify type of data and performance envelope. If more than one task or task sequence is selected on one page, use a corresponding number of asterisks.

As you read each task ask yourself

(1) Is this a task where the operator makes an input to the system?

If the answer is no, proceed to the next task.

If the answer is yes, ask yourself.

(2) Could some kind of data be obtained after this task is accomplished that would provide some indication of the operator's performance or performance level?

If the answer is no, proceed to the next task.

If the answer is yes, please indicate in the far right hand column of each page.

(3) What kind of data could be obtained or should be obtained to determine operator performance with respect to system design?

(4) Could the data that is obtained be used to construct a performance envelope around the particular task in question?

If the answer is no, proceed to the next task.

If the answer is yes, please indicate

(5) What limits, constraints, tolerances, accuracies or other criteria could be used to define a performance envelope for a given task?

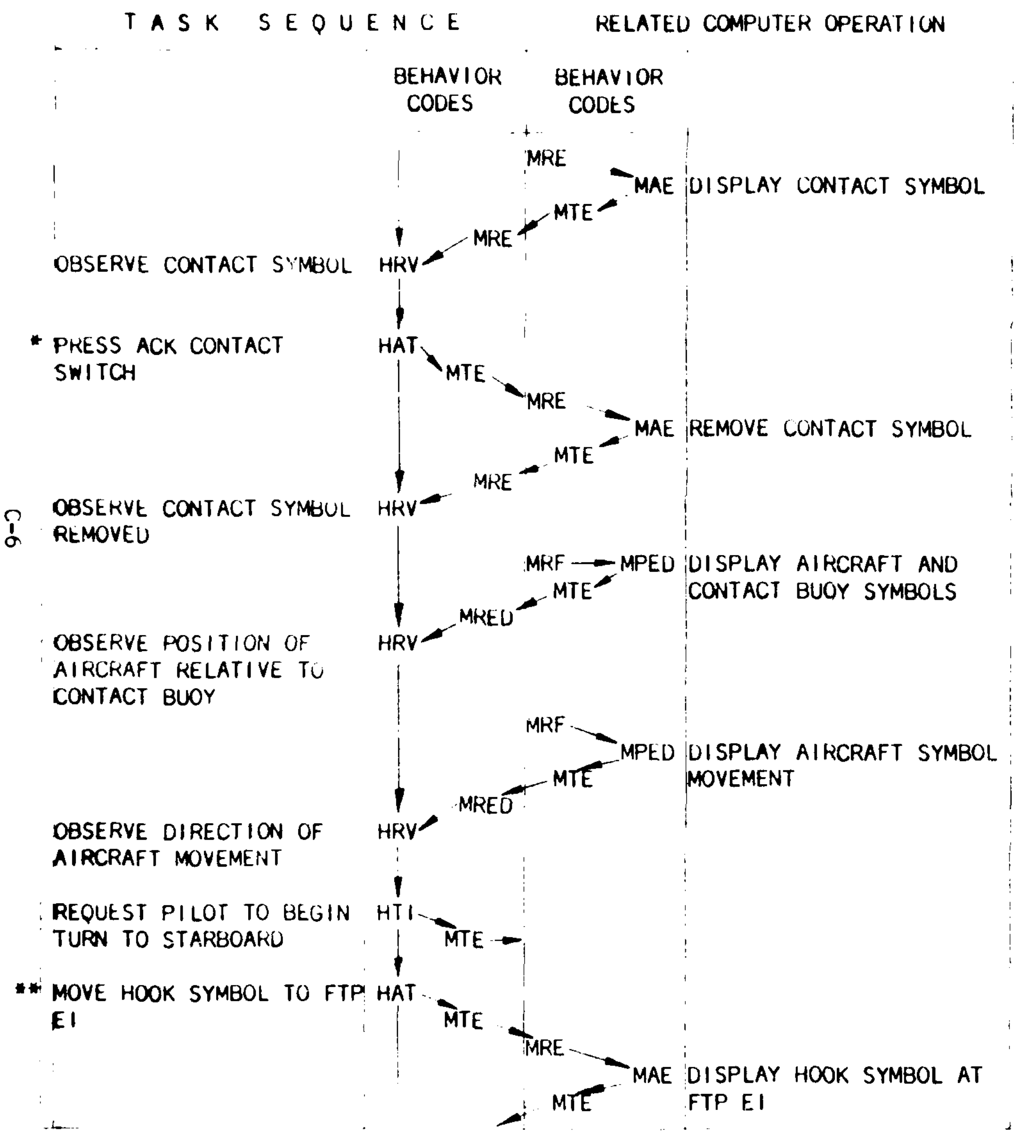
For example, the data may concern how accurate the operator's inputs should be. Then the question would be what is the level of accuracy that would be the difference between acceptable and unacceptable.

CONSTRUCTION OF BEHAVIORAL CODE WORDS

FIRST LETTER		
HUMAN OR MACHINE	H	(Human)
	M	(Machine)
SECOND LETTER		
BEHAVIOR	A	(Act)
	D	(Decide)
	T	(Transmit)
	R	(Receive)
	S	(Store)
	P	(Use Previously Stored Information)
	M	(Monitor)
THIRD LETTER		
MEANS OF PERFORMANCE	S	(Speech)
	P	(Phone, Sound-power)
	I	(Intercom)
	E	(Electrical or Electronic)
	T	(Touch)
	V	(Visual)
	F	(Filed)
FOURTH LETTER		
DISPLAYED OR NOT	D	(Displayed)
	Blank	(Not Displayed)
FIFTH LETTER		
INVERSE MEANING	G	(Go, yes, normal, etc.)
	N	(No-Go, no, abnormal, etc.)

SYSTEM - P-3C ASW AIRCRAFT
OPERATOR - TACCO
PHASE - SEARCH
EVENT - INITIAL CONTACT

PAGE 1 OF 6
REVISION NO 0
DATE 14 JUN 1971



LIST OF TYPE OF DATA AND CRITERIA FOR EVALUATING DATA

* DATA -
1. Total number of actuations.
2. Number of correct actuations.

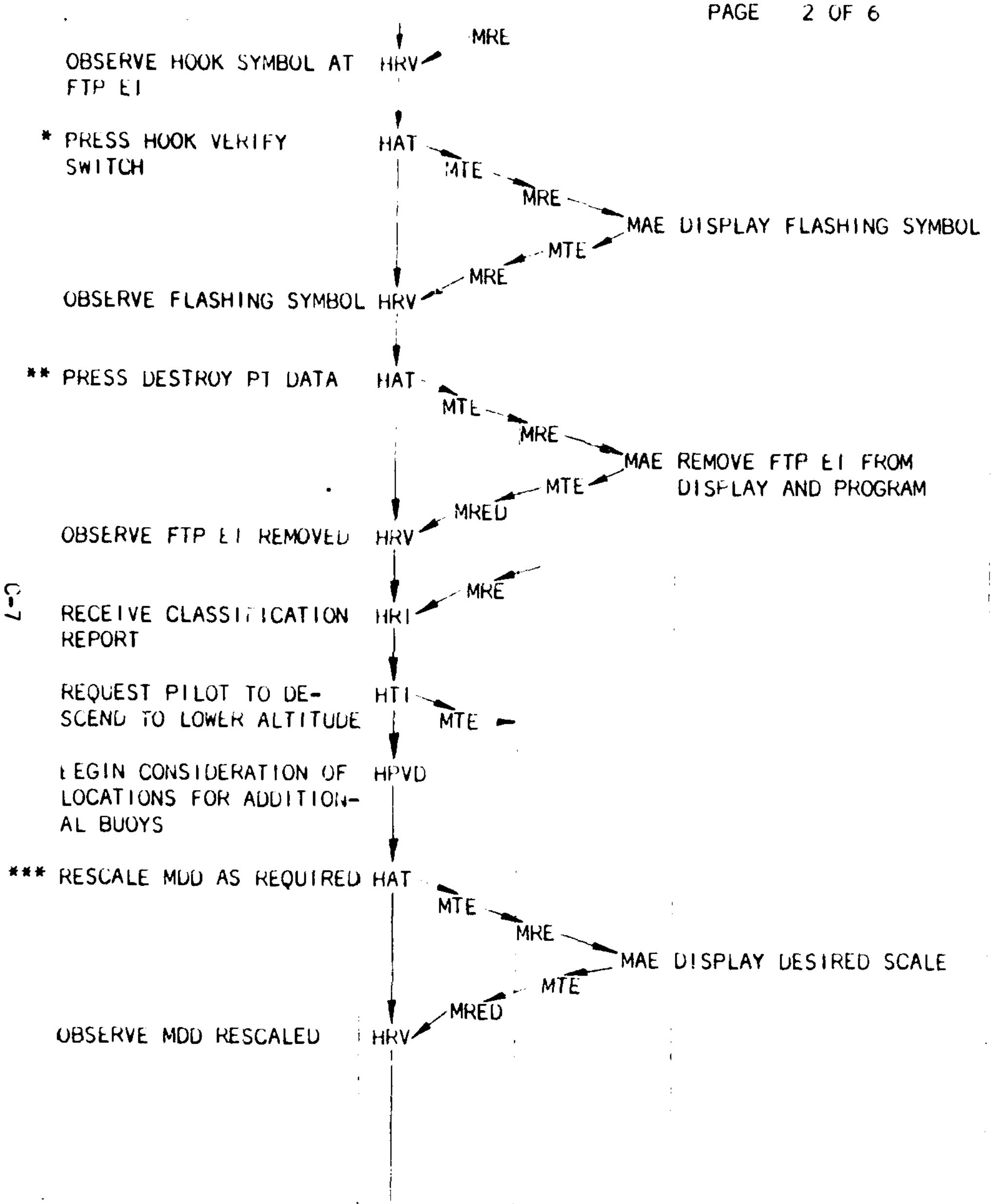
CRITERIA -
1. 100% Correct actuations.

EVALUATION -
1. Percentage of correct actuations.

** DATA -
1. Total number of times that FTP symbols are intended to be hooked.
2. Number of times that FTP symbols are actually hooked.

CRITERIA -
1. 100% of times that FTP symbols are intended to be hooked.

EVALUATION -
1. Ratio of actual to intended.



LIST OF TYPE OF DATA AND
CRITERIA FOR EVALUATING DATA

* DATA -

1. Total number of actuations.
2. Number of correct actuations.

CRITERIA -

1. 100% Correct actuations

EVALUATION -

1. Percentage of correct actuations

** DATA -

1. Total number of times that FTP symbols are intended to be hooked.
2. Number of times that FTP symbols are actually hooked.

CRITERIA -

1. 100% of times that FTP symbols are intended to be hooked.

EVALUATION

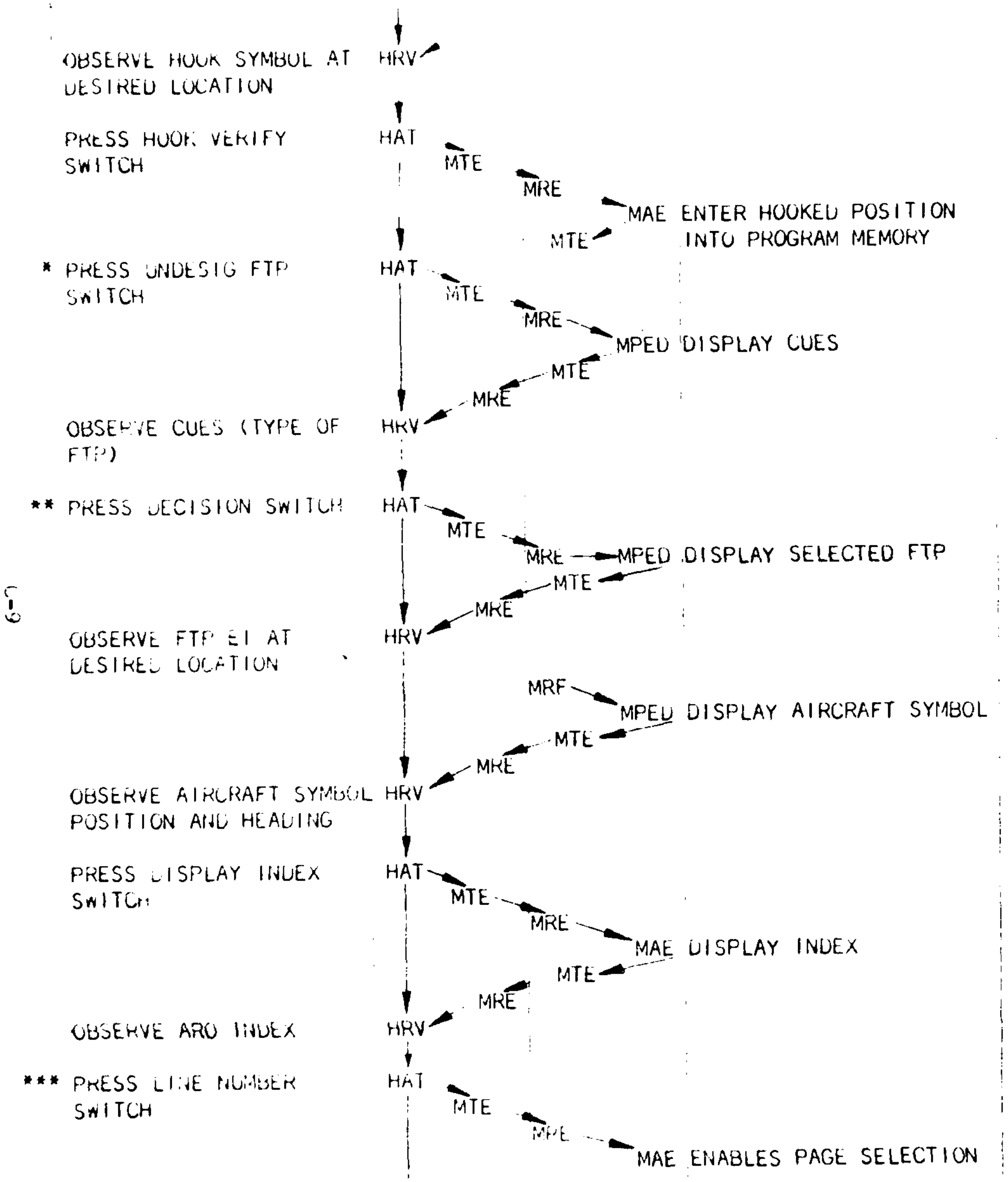
1. Ratio of actual to intended.

SYSTEM - P-3C ASW AIRCRAFT
OPERATOR - TACCO
PHASE - SEARCH
EVENT - REFINE SEARCH PATTERN

PAGE 3 OF 6
REVISION NO 0
DATE 14 JUN 1971

LIST TYPE OF DATA AND
CRITERIA FOR EVALUATION OF DATA

TASK SEQUENCE	BEHAVIOR CODES		RELATED COMPUTER OPERATION BEHAVIOR CODES	RELATED COMPUTER OPERATION
			MRF IMAE MTE	DISPLAY BEARING LINE FROM CONTACT BUOY
OBSERVE BEARING LINE FROM CONTACT BUOY	HRV	MRE		
			MRF MPED MTE	DISPLAY BEARING LINE AND AIRCRAFT POSITION AND HEADING
COMPARE POSITION OF BEARING LINE TO AIRCRAFT POSITION/HEADING	HRV	MRE		
CONSIDER WATER CONDITIONS	HA			
			MRF MPED MTE	DISPLAY RADAR CONTACTS
CONSIDER RADAR CONTACTS	HRV	MRE		
CONSIDER ALL OTHER AVAILABLE TACTICAL INFORMATION	HPVD			
DETERMINE LOCATION FOR FIRST PATTERN REFINEMENT BUOY	HD			
MOVE HOOK SYMBOL TO LOCATION FOR FIRST PATTERN REFINEMENT BUOY	HAT	MTE MRE	MAE MTE	DISPLAY HOOK SYMBOL AT DESIRED LOCATION
		MRE		



* Same as Press ACK CONTACT Switch
Page 1 of 6 (C-6)

** DATA -
1. Total number of times that a Decision switch is pressed.
2. Number of times that the correct Decision switch is pressed.

CRITERIA -
1. 100% Correct actuations.

EVALUATION -
1. Percentage of correct actuations.

*** DATA -
1. Total number of actuations.
2. Number of correct actuations.

CRITERIA -
1. 100% Correct actuations

EVALUATION -
1. Percentage of correct actuations.

SYSTEM P-3C ASW AIRCRAFT
OPERATOR - TACCO
PHASE SEARCH
EVENT REFINE SEARCH PATTERN

TASK SEQUENCE	BEHAVIOR CODES	BEHAVIOR CODES	RELATED COMPUTER OPERATION
* ENTER LINE NUMBER FOR NAVIGATION DATE	HAT → MTE →	MRE → MAE	STORE LINE NUMBER IN PROGRAM MEMORY
** PRESS EOM SWITCH	HAT → MTE →	MRE → MAE	DISPLAY NAVIGATION DATA
	MRE ←	← MTE	
OBSERVE NAVIGATION DATA	HRV		
		MRE → MAE	DISPLAY CONTACT SYMBOL
	MRE ←	← MTE	
OBSERVE SECOND CONTACT SYMBOL	HRV		
*** PRESS ACK CONTACT SWITCH	HAT → MTE →	MRE → MAE	REMOVE CONTACT SYMBOL
	MRE ←	← MTE	
OBSERVE CONTACT SYMBOL REMOVED	HRV		
REQUEST BEARING UPDATE ON FIRST CONTACT BUOY	HTI → MTE →		
	HAT → MTE →	MRE → MAE	DISPLAY CUES
	MRE ←	← MTE	

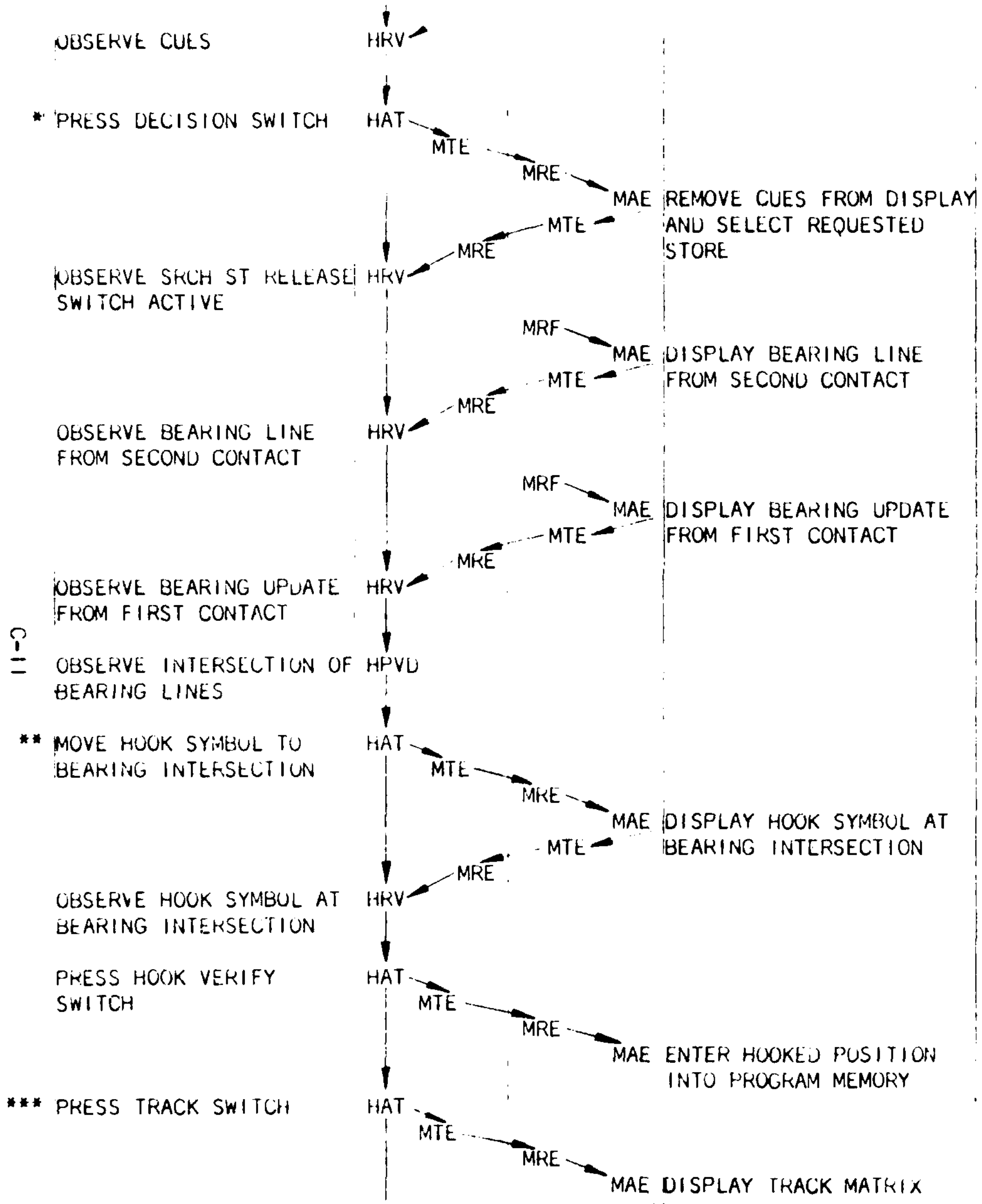
LIST TYPE OF DATA AND
CRITERIA FOR EVALUATION OF DATA

* DATA -
1. Total number of times that an Index Line Number is selected.
2. Number of times that the correct Index Line Number is selected.

CRITERIA -
1. 100% Correct selection.

EVALUATION -
1. Percentage of correct selections.

** See Press LINE NUMBER switch on page 4 of 6 (C-9).



* Same as PRESS ACK CONTACT SWITCH, Page 1 of 6 (C-6).

** DATA -
   1. Miss distance.

CRITERIA -
   1. Zero miss distance.

EVALUATION -
   1. Calculate mean and standard deviation.

*** See Press ACK CONTACT Switch on page 1.